

Firms as Tax Collectors*

Pablo Garriga[†]

Job Market Paper

Dario Tortarolo[‡]

This version: November 12, 2021
[Most recent version here](#)

Abstract

We study the implications of delegating tax collection duties on firms. We exploit a major reform to the withholding regime of the Turnover Tax in the City of Buenos Aires, where a substantial number of large firms were appointed to act as Collection Agents (CAs) based on a predetermined revenue threshold. Combining rich firm-to-firm administrative data along with quasi-experimental methods, we show that: (i) firms newly appointed as CAs do not experience any effects on their reported business activity; (ii) firms with pre-existing commercial ties to CAs increase their self-reported income; (iii) the government collects more tax revenue. Additionally, the analysis of a subsequent reform that reduced third-party collection of the tax shows that firms respond symmetrically by reducing their reported income. These results are in line with other papers suggesting that reforms to the tax administration can have considerable impact when it comes to raising revenue and building tax capacity. In particular, our findings can provide guidance to other middle- and low-income countries on ways to determine who the right tax collector is as a function of the level of development.

*We would like to thank John Friedman, Kenneth Chay and Pierre Bachas for research guidance and very fruitful discussions. We are also thankful to Anne Brockmeyer, Daniel Björkegren, attendees to Brown University Applied Micro Lunch, NBER Business Taxation in a Federal System, for their insightful comments and suggestions. We are grateful to Hilario Iñiguez and Octavio Pera from AGIP for helping us to understand the data and for making this work possible. The views expressed are those of the authors alone and do not reflect the views of AGIP.

[†]Brown University, pablo_garriga@brown.edu

[‡]University of Nottingham & IFS, dario.tortarolo@nottingham.ac.uk

1 Introduction

Governments in developing countries struggle to raise revenue and build tax capacity. Increasing taxes and mitigating non-compliance are the conventional direct practices considered by authorities. Yet, achieving large-scale capacity oftentimes requires fundamental transformations of tax administrations in the longer run.¹ How taxes are collected is usually at the forefront of tax authorities' debates. Among the potential instruments, withholding systems—where third-parties collect and remit taxes owed by related parties—have historically played a central role in easing tax administrations.

Tax withholding systems are ubiquitous. In the case of the income tax, part of its global success is due to employers withholding at source from their employees (Besley and Persson, 2014). Similarly for the VAT, its built-in withholding mechanism has been shown to enhance compliance (Pomeranz, 2015; Waseem, 2020). Withholding of business income is particularly common in developing countries where governments lack the resources and capacity to accurately measure and tax firm activity (Slemrod, 2008; Brockmeyer and Hernandez, 2019). In this case, financial institutions and large companies collect taxes owed by other firms in their commercial network. The use of such collection device for indirect taxes has surged over the last decade (Figure 1), but little is known about its implications.

This paper asks what are the effects of delegating tax collection duties on firms. For this purpose we analyze an unprecedented expansion of a withholding scheme to collect the Turnover Tax in the City of Buenos Aires, Argentina. Withholding implies two main changes on tax collection: *who* remits the tax and *when* it is payed. Under the standard filing procedure, firms are required to report their monthly sales and then apply the corresponding tax rate to it. Under the withholding scheme some firms are appointed to act as Collection Agents (CAs) and collect part of the tax in advance from their commercial partners whenever there is a purchase or a sale. While in a first-best world the point of collection of the tax would be irrelevant, it has been shown that in practice changes in the collection mechanism matter for compliance (see Slemrod, 2008; Kopczuk, Marion, et al., 2016).

Our goal is to provide insights on the implications of withholding for both the collecting party and linked taxpayers. For CAs, withholding is an administrative burden as they now have to declare and remit taxes on behalf of others. It also implies that they are under closer scrutiny from the government who can penalize them if they fail to fulfill their

¹Such administrative reforms have proven highly effective according to recent research in Indonesia by Basri et al. (2019).

role. Finally, withholding allows these firms to get a hold of additional funds that may be pocketed in the time between collection and remittance. For linked firms withholding generates third-party information which may alter their enforcement perceptions. Also, the withheld amount itself may act as a lower bound on self-reported sales and tax owed. Lastly, it might distort the choice of trade partners towards non-CAs in order to minimize the tax burden.

We bring new evidence on this topic by exploiting a reform in 2016 that implied a major expansion of the withholding scheme set up to collect the Turnover Tax. Under this reform firms were appointed to act as CAs according to a rule based on size: firms whose gross income (sales) in the preceding year was greater than AR\$ 60 million (located at the 97th percentile of the revenue distribution) were automatically appointed as CAs. Appointed firms then started to withhold the tax from their commercial partners. This implied an increase in the share of taxpayers' tax liability collected indirectly at source by CAs *in lieu* of direct payments to the tax authority. The setting allows us to estimate the *direct effects* of expanding the withholding scheme on the firms who are put to the task of collecting taxes, and the *indirect effects* on those who now face an increase in the amount of taxes being withheld at source. To our knowledge, this is the first paper that is able to document the full roll-out of a policy of this kind—papers analyzing withholding of business income taxes omit studying the effects on the withholding agents who actually bear the compliance costs.

Another novel feature of our paper is the data used for the empirical analysis. We rely on two administrative records processed by the Tax Administration. First, monthly Turnover Tax declarations for nearly the universe of firms operating in the City of Buenos Aires. Second, transaction-level details of purchases and sales provided by firms acting as CAs. These records allow us to trace out the commercial linkages between firms, a key feature for our analysis. By combining the two data sources we obtain a panel of 250 thousand firms and 1.6 million client-supplier pairs spanning from 2015 to 2020. We combine the data coupled with quasi-experimental methods that leverage on the exogenous variation created by the reform to estimate these effects.

Our empirical analysis begins by documenting the implementation of the program and the consequent expansion of the withholding net to provide some insights on how the policy actually worked in practice. We show non-parametric graphical evidence suggesting that newly appointed firms quickly took up their roles and started collecting the tax very much like other pre-existing CAs. We then proceed to study the direct effect of the reform by trying to determine whether appointing firms as CAs affects their economic

activity. We use a regression discontinuity design that exploits the sharp change in the probability to work as a CA instituted by the 2016 reform. While previous enrollments to the CA regime had been determined on a case-by-case basis, the November 2016 appointments were decided by applying the “AR\$ 60 million” cutoff rule on 2015 sales. We show evidence that the rule was followed closely, but with some exceptions: firms belonging to some specific industries were either left out or included in the collection scheme regardless of their 2015 sales. Taking these features into account, we estimate whether CAs change their behavior by comparing reported income for firms close to the cutoff. The evidence suggests that CAs did not experience any effects on their business activity. We attribute this result to the fact that appointed firms were among the largest—and presumably most formal—firms in the economy, with highly streamlined tax filing practices, such that collecting taxes from partners and remitting them to the government does not imply an increased burden or a change in enforcement perceptions.

Next, we analyze the response of taxpayers to withholding. We focus on firms who regularly trade with newly-appointed CAs and, as a consequence of the reform, experience an increase in the amount of tax collected through the withholding mechanism. We use a differences-in-differences methodology that relies on the fact that some firms were more exposed to the reform than others. Exposure is defined by the presence of new CAs in the network of commercial partners of a given firm: firms in the control group are only linked to pre-existing CAs and do not experience any change in the way their tax is collected; on the other hand, firms in the treatment group are linked to newly appointed CAs, implying that part of the tax that was payed through direct payments is replaced by withholding after the reform is in place. The first-stage results show that both the number of linked CAs and the share of due taxes withheld at source increase sharply among firms more exposed to new CAs, confirming that the reform only had bite among the treated group. Moving on to the analysis of behavioral responses, our findings suggest that firms in the treatment group increased their reported gross income by 5 percentage points relative to the control group. A potential interpretation of this result is that withheld taxes act as a lower bound on self-reported income: if firms were under-reporting their sales before the reform, having a larger part of the tax remitted by a CA might have forced them to increase the amount they must report.

In the last part of the analysis, we leverage a different policy change that took place in September 2018 and affected a subgroup of firms linked to CAs in the opposite direction. This allows to shed light on whether firms respond symmetrically to a *decrease*—rather than an increase—in withholding by reporting fewer sales. To this end, we analyze a tem-

porary *reduction* to the withholding on financial transactions that was targeted to firms whose sales in 2017 were below AR\\$ 10 million. While those below the threshold were waived from being withheld by banks on any deposit received or any transfer made to other firms, firms above the threshold remained subject to bank withholding. Using a difference-in-discontinuity design, we show that firms below the cutoff effectively experienced a reduction in the share of taxes withheld at source and responded by reducing their reported sales. This result strikes us as remarkable and could suggest that a reduction in the coverage of third-party information gives firms more freedom to misreport their sales and thus to reduce their tax liability. In other words, the same set of firms who reported more sales when they became subject to withholding from CAs in 2016, exhibit a reduction in their sales when the bank withholding decreases in 2018.

Taken together, our findings suggest that withholding by large firms can be an effective tax collection tool. On the one hand, appointing firms to collect taxes does not seem to hurt their activity directly. On the other hand, substituting direct tax payments for withholding at source increases reported sales of firms linked to newly appointed CAs which, in turn, increases total tax revenue.

This paper builds on several strands of literature. First, it speaks to the literature on tax compliance and enforcement that studies the drivers behind evasion and avoidance behavior (Allingham and Sandmo, 1972; Kleven, Knudsen, et al., 2011; Slemrod and Yitzhaki, 2002; Slemrod, 2019), particularly to those stressing two critical aspects of modern taxation: how firms play a key role as fiscal intermediaries and how tax compliance critically depends on automatic features such as withholding (Kleven, Kreiner, et al., 2016; Slemrod, 2008).

We are closely related to two empirical papers describing the implementation of withholding mechanisms, Brockmeyer and Hernandez (2019) and Waseem (2020). These papers provide compelling evidence on how the “remittance irrelevance” proposition—whereby the identity of the remitting party does not matter for tax collection purposes—is very unlikely to hold in a developing country. Brockmeyer and Hernandez (2019) analyze withholding by credit-card companies in Costa Rica. They exploit variation in the coverage by third-party reporting and in the withholding rate to show that withholding increases sales tax revenue. The mechanisms behind this are default payment and increased enforcement perceptions. Waseem (2020) studies the roll out of the VAT in Pakistan. Manufacturing firms, already in the tax net, increase their reported revenue when the tax is extended upstream to the energy sector and their inputs become subject to withholding. This behavior is rationalized by the withheld amount acting as a lower bound on self-

reported sales. Our setting is more general since withholding duties were assigned in a manner that was close to exogenous—for CAs themselves and for linked firms, who became more exposed to third-party collection and remittance. Additionally, the reform we analyze was broader-ranging as it was targeted to the whole economy, not a specific sector. These features allow us to determine more precisely in which cases might this tax collection mechanism be more effective.

We believe that some of the features of our setting can provide important takeaways for other middle- and low-income countries. In this sense, our paper also contributes to the literature on taxation and development whose goal is to determine what is the optimal tax administration intervention as a function of development (Keen and Slemrod, 2017; Best et al., 2015; Bergeron et al., 2021; Basri et al., 2019). The reduction in the barriers to implement more sophisticated tax administration systems means that the policy studied in this paper will become accessible for transitioning low-income countries. We think that our analysis can provide insights on ways to determine the optimal set of tax collectors are in each setting.

The paper is organized as follows. Section 2 provides details on the tax, the reform that we study, and some suggestive evidence of its macroeconomic implications. Section 3 specifies the data sources used in our analysis. Section 4 describes the conceptual framework used to rationalize the empirical results. Section 5 documents the expansion of the withholding scheme and its direct effect on CAs while Sections 6 and 7 study the indirect effect on taxpayers. Section 8 concludes.

2 Context

The Turnover Tax (TT) is a sub-national tax levied on firm revenue with no deductions for costs. It applies to all transactions taking place in the supply chain—i.e., business to business and business to consumer transactions. This implies that it creates “cascading effects” where final goods are taxed multiple times throughout production. Despite its distortive effects, the tax is simple to collect which explains why it is used in all of the 24 jurisdictions in Argentina.² In particular, for the City of Buenos Aires, the TT constitutes the main source of revenue accounting for about 75% of total tax collected. Although firms

²In fact, Hansen et al. (2021) shows that the tax is re-emerging in the US, where nine states are currently using it. They attribute this to: (i) the large tax base (TT taxes services while Retail Sales Taxes do not) that can generate large revenues with low rates, (ii) its main drawback—taxing inputs—also occurs in the case of the RSTs, and (iii) its broad tax base limits firms tax evasion opportunities.

must file their returns on a monthly basis, the city also uses a withholding scheme that relies on firms to collect part of the tax in advance from their commercial partners—going forward, we will refer to these two alternate collection mechanisms as “direct payment” and “withholding” mechanisms. In 2021, 66% of TT revenue was remitted by withholding firms and the remaining 34% was collected through direct payments of taxpayers.

In practice, there are two distinct withholding mechanisms depending on the identity of the collecting party: the CA can either be a commercial partner or a financial institution (banks, credit- or debit- card companies). In 2021, the former contributed with approximately 30% of TT revenue while the latter collected about 25% of TT revenue (20% by banks and 5% by cards). Our main focus throughout the paper will be on withholding by commercial partners. However, we will also analyze a policy reform that temporarily eliminated the withholding by banks on a subset of firms.³ We will provide more details on how the tax collection mechanisms work in practice in Section 4.⁴

2.1 The reform

Historically, tax collection in the City of Buenos Aires relied mostly on direct payments. Withholding by CAs was introduced in 1985; between 1985 and 2016 some large firms were appointed to act as CAs on a case-by-case basis. These firms would withhold part of the taxes owed by their commercial partners whenever there was a purchase or sale.

In July 2016, a tax resolution was passed with the intention of expanding the net of firms acting as tax collectors.⁵ It established a new appointment criterion based on pre-determined sales: firms whose gross income in 2015 was greater than or equal to AR\$ 60 million would be automatically enrolled as CAs. Appointed firms were notified in October and started their collection duties in November 2016. Firms appointed before this norm was passed remained in the tax collection scheme regardless of their revenue. In the years 2017 through 2020 there were no new appointments. For clarity, we summarize the timeline of these events below.

The direct implication of the reform was a sharp increase in the number of CAs. Figure 3 plots the number of firms acting as CAs across time. Its pattern exactly matches the timeline described above. For the period prior to it, there were small increases in the number of active CAs—in fact, between 1985 and 2001 there was only one CA, the tax ad-

³Withholding can also take place at Customs, but this channel only contributes a small fraction of the total tax collected.

⁴For more details about the Turnover Tax and the way it is collected see Appendix A.

⁵Resolutions 364/AGIP/2016 and 421/AGIP/2016.



ministration itself. Between October 2016 and November 2016 the stock of CAs more than doubled with the implementation of the new appointment rule. Afterwards, the number of active CAs displayed a gradual decrease as some firms exited.

Macroeconomic evidence. The effects of the reform were considerable and immediate. Right after being appointed, CAs started to withhold the tax from their commercial partners. This implied an increase in the share of taxpayers' tax liability collected indirectly at source by CAs *in lieu* of direct payments. Appendix Figure C.4a shows that tax collected through withholding increased from 30% to 45% around the time of the reform.

The reform had a noticeable impact on revenue as well. To show this effect, we compare tax revenue in the City of Buenos Aires relative to the Province of Buenos Aires. These are the largest jurisdictions in the country in terms of economic activity and their economic performance track each other closely, making each one a natural benchmark of the other. Appendix Figure C.4c shows the percentage difference in tax revenue of the City of Buenos Aires relative to the Province, taking the month prior to the reform as baseline period. During the periods prior to the reform the City was performing worse than the Province yet, right at the time of the reform, there is a trend break after which the outcome becomes more favorable for the City. The macro evidence shows that the reform had a large impact. Our goal in the remainder of the paper will be to flesh out the microeconomic components.

3 Data

Our analysis combines two sources of data from the Tax Administration Office in the City of Buenos Aires (AGIP). The first data source consists of monthly tax declarations for nearly the universe of firms operating in the City of Buenos Aires for 2015-2020. These declarations are filed electronically through an account each taxpayer has in the tax administration's website. We observe all line items completed in a standard filing procedure (gross income, amount of tax withheld distinguishing sales and purchases, outstanding credits or debits) as well as a series of firm characteristics used to determine the corre-

sponding tax and withholding rates (firm type: small or large; sector: 995 six-digit codes; location: local or out-of-province).

To construct our second data source, we leverage on a unique feature of our setting: firms acting as CAs are required to file a supplementary “invoice summary” that contains transaction-level details of purchases and sales that occurred within a given month. The line items included in this form are: total value of transaction, amount withheld, distinction between purchase or sale. We use these filings to trace-out commercial linkages between firms, allowing us to establish the relationships that exist between CAs and their commercial partners. A limitation of this data is that trade linkages are only observed when a firm becomes CA—that is, when they start filing the supplementary form that reveals transactions. This implies that for some firms we observe linkages since the beginning of our sample, in 2015, while those linked to CAs appointed with the reform, we observe linkages only at, or after, November 2016. Figure 4 illustrates the process by which we construct our full dataset and further stresses the sequence in which the information is revealed. Once the two sources of data are combined, we obtain a panel of 250 thousand firms and 1.6 million client-supplier pairs spanning from 2015 to 2020.

Table 1 presents summary statistics for the period prior to the 2016 reform. Panel A describes the full sample. It contains over 190,000 thousand firms, roughly 5 percent of them are CAs. Overall, the firm size distribution in terms of revenue is right-skewed. CAs are orders of magnitude larger than regular firms which corroborates the enrollment heuristics followed by the government of enrolling large firms, even before the 2016 reform. Panel B corresponds to the estimating sample used in Sections 6 and 7. When evaluating the response of taxpayers we restrict the sample of firms to those that file taxes regularly (we drop firms that report positive sales in less than 50 percent of the periods), as a direct implication this subset of firms is larger than that the full sample.

4 Conceptual framework

In this section we focus on how tax collection works in practice and provide some insights on how changing the tax collection mechanism might affect firm behavior. Figure 2 summarizes the main features of the direct payment and withholding mechanisms in a simplified setting with four agents: supplier, retailer, consumer and tax administration. The supplier sells inputs to the retailer who pays $\$Y$ for the goods. In turn, the retailer takes the inputs and turns them into final goods which are sold to the consumer who pays $\$X$ for them. As described earlier, all firms along the supply chain are liable for the tax.

In our setting the supplier and the retailer respectively face tax rates τ_S and τ_R while the withholding rate (which is a fraction of τ) is represented by $\alpha \in (0, 1)$.

Direct payment. Under the regular payment mechanism, taxpayers self-report their sales at the time of payment. The supplier reports \$X and the total tax liability for the period is $\tau_S X$. Similarly for the retailer, it reports \$Y, such that liability is $\tau_R Y$.

Withholding. Under the withholding scheme firms acting as CAs are in charge of withholding a part of the total tax liability from their commercial partners and remitting this amount to the tax administration. This task is performed in addition to their own duties as taxpayers—i.e., collection and remittance of third-party taxes runs in parallel to fulfilling one's own tax payments.

Withholding may be performed by either the seller or the buyer in a transaction. When the seller acts as CA it adds the withheld tax to the total sale amount. In our setting, this is represented by the supplier charging an extra amount for the sale to the retailer, $X(1 + \alpha\tau_R)$, which is then remitted to the tax administration (along with their own taxes). When the retailer's tax is due it will only pay the outstanding balance that remains after deducting the remitted amount from the total tax liability, $\tau_R Y - \alpha\tau_R X$. When the buyer acts as CA it subtracts the withheld tax to the total purchase amount. In our setting, the retailer pays $X(1 - \alpha\tau_S)$ to the supplier and remits it to the tax administration where the funds become available to the supplier to deduct from its due taxes.

Implications. Withholding through CAs implies two main changes on tax collection: *who* remits and *when* the tax is payed. The direct payment mechanism implies that the tax is filed by the liable party on the due date. On the other hand, the withholding mechanism implies that the tax is payed by a third-party in advance—the tax is withheld at source and remitted on a regular basis. These changes carry associated implications for both the collecting party and the taxpayer.

For CAs, appointment leads to closer scrutiny and increased penalties if they do not fulfill their role as CAs. Thus, the appointment event itself might trigger changes in enforcement perceptions that lead to an increase in reported sales for firms who were previously underreporting (see Brockmeyer and Hernandez, 2019).

Additionally, withholding is an administrative burden as it requires CAs to keep a registry of transactions with trade partners and to file an additional form detailing such transactions. As pointed out by Slemrod (2008), this burden might be offset by the “cash-flow benefit” of withholding: CAs may pocket the funds they withhold in the time between collection and remittance, equivalent to an interest-free loan. However, the magni-

tude of these two contrasting forces is likely to be small in our setting. On one hand, the collection process is highly streamlined: when a firm is appointed CA it is provided a tax collection software that automatically incorporates withholding to its transactions. On the other hand, firms must transfer the withheld funds every month following a pre-specified schedule only allowing for a small window of time between the remittance deadline and the tax triggering-event where they could dispose of the funds.

For linked firms there are three direct implications of switching to the withholding collection mechanism. Withholding enables each business-to-business transaction to be recorded at two places, generating third-party information. This may alter enforcement perceptions since the information could be used by the administration to detect non-complying firms. In addition, the withheld amount itself may act as a lower bound on self-reported sales and tax owed as building up tax credit might raise a red flag. Finally, it might distort the choice of trade partners towards non-CAs: if withholding implies that a firm subject to it ends up facing a higher effective tax rate, the firm may decide to switch trade partners in order to minimize its tax burden.

We can derive testable predictions for the first two implications using the framework of firm tax evasion under third party reporting developed in Carrillo et al. (2017), Brockmeyer and Hernandez (2019), and Waseem (2020). Note that these papers usually consider settings where there is a VAT or sales tax allowing for deduction of costs. Our setting is slightly less complex since the TT does not allow for any type of deduction. Consider a firm that is already subject to some amount of third party withholding. If evasion costs are sufficiently low it will underreport sales up to the point that tax owed is completely offset by withheld funds. Now, when a trade partner is appointed CA and additional transactions become subject to withholding, the firm can only avoid falling in the negative liability territory by increasing reported sales. Therefore, we should expect to see an increase in reported sales for firms more exposed to withholding as some of them will try to avoid raising a red flags in the tax administration. We should also expect to see an upward shift in the distribution of tax liability as firms try to bunch close to the lower bound set by withheld funds. Finally, as shown in Pomeranz (2015), we would expect the response of reported sales to be stronger for certain sectors where informality is high or at the retail stage where the cost of evasion is relatively lower since these sectors lack “preventive deterrence”.

Regarding the third implication, Gadenne et al. (2020) show that supply chain distortions might occur as firms who were previously underreporting face an increase in the tax burden when their current supplier becomes CA. Firms who can choose among several al-

ternative suppliers will face lower input price if they switch to one who is not a CA. Thus, we would expect some trade links to break or become less frequent, or show a decrease in trade volume.

5 Expansion of the withholding scheme

We begin the empirical analysis by documenting the expansion of the withholding scheme. Specifically, we describe the implementation of the appointment rule, the expansion of the withholding net and the response of newly-appointed CAs as they take on their new role.

We use a series figures to provide non-parametric graphical evidence of whether the appointment rule worked in practice and how quickly the new CAs started withholding from linked firms: take-up was high and firms responded quickly. Next, we analyze the response of appointed firms. Here, we use a regression discontinuity design (RD) that exploits the sharp change in the probability to work as a CA based on 2015 gross income (pre-reform) to test whether CAs change their behavior by comparing reported income for firms close to the cutoff. Our results suggests that CAs did not experience any effects on their business activity, we interpret this as evidence that the policy did not imply a significant change in their business structure.

Implementation of the appointment rule. Figure 5 shows the 2015 gross income distribution for the whole sample of firms. The figure reveals an important fact about the type of firms that were targeted by the reform: they were among the largest in the economy. The cutoff roughly corresponds to the 97th percentile of the firm size distribution. The bottom panel zooms in around the cutoff: there are no discontinuities around the 60 million threshold which confirms that firms were not able to manipulate the running variable used to determine appointment into the CA program. This is a promising fact that suggests that we can use firms around the limit to identify causal effect.

According to the appointment rule firms above the AR\$ 60 million threshold should have been appointed to act as CA. Figure 6 provides evidence of how binding this rule actually was. On the x axis we take the running variable and split it into bins of size AR\$ 10 million, the y axis is the probability of Treatment (i.e., being appointed CA). Under perfect compliance, the probability of treatment should be 0 for firms below the threshold and 1 for firms above.

Expansion of the withholding net. Non-compliance of the duties assigned to CAs is heavily penalized. However, these regulations by themselves do not ensure that firms will take

up their role as tax collectors. The next figures are intended to show that appointed firms rapidly took up their new duties. To do so, we compare their behavior with a group of pre-existing CA appointed in 2012. We use this group as a benchmark of what we would expect to be a “normal” behavior from firms who are already established in their roles as CAs. Figure 7a analyzes the extensive margin of collection duties: for each month it shows the proportion of CAs in each group that are withholding from commercial partners. It takes roughly a year for new CAs to fully take on their role. However, their starting point is very high: on the 1st month of appointment, close to 70% of new CAs are withholding from other firms. Figure 7b in turn, shows the intensive margin of collection duties: for each month it shows the average number of firms that get withheld in each group. New CAs on average withhold from fewer firms, this might be due to the fact that they’re smaller/younger than old CAs. Taken together, all these elements provide evidence that the policy worked in practice. The final step in our description of the expansion of the withholding scheme is to know whether it had any direct effect on newly-appointed CAs.

5.1 Response of newly-appointed CAs

We exploit the sharp change in the probability of being enrolled as CA established by the appointment rule in our estimation of the direct effects of the expansion of the withholding scheme. We use an RD framework to compare changes in gross income (sales) for firms close to the AR\$ 60 million cutoff. We expect to see if any of the predictions that were presented in Section 4 can be confirmed in the data.

Figure 8 and Table 2 show the results for the RD. Our evidence suggests that CAs did not experience any effects on their business activity. We attribute this result to the fact that appointed firms were among the largest—and presumably most formal—firms in the economy, so it makes sense that tax collection duties did not imply a significant compliance costs to them, nor did they change their perception on tax enforcement.

If the firms who are becoming CAs are themselves having a large change in their business structure then that would be a concern for the analysis that follows because it would mean that its not just becoming a CA but something fundamental bout this business interaction is changing. The results shown in this section suggest that firms that became CAs did not change their behavior. In the remainder of the paper we focus on the impact on business partners.

6 How do firms respond to an increase in withholding?

In this section we study the indirect implications of expanding the withholding scheme used to collect the tax. In other words, we focus on the response of taxpayers to an increase in the amount of tax withheld from them. The ideal experiment to answer this question would be to randomly allocate CAs throughout the economy and compare the differential response of firms who have commercial linkages with these CAs relative to those who do not. While we cannot randomly assign tax collection duties across firms, we can leverage on the design of the reform and use the built-in variation in exposure to CAs to provide a causal effect of withholding on compliance. More specifically, we rely on the following quasi-experiment. The expansion of the withholding scheme implied that some large firms were appointed as CAs and, as shown above, this lead to an exogenous and sharp increase in the number of CAs. This had no immediate consequences for firms not directly targeted by the reform—i.e., regular taxpayers. However, some saw a larger increase in the number of CAs withholding from them as a by-product of their commercial linkages. This change in the number of CAs was arguably exogenous to them as it is unlikely that their choice of commercial partners would have been related to the size threshold used by the appointment rule.

The key econometric challenge that must be tackled to precisely estimate the effect of withholding on compliance is to distinguish it from other time-varying shocks that might have occurred around the time of the reform. To address this issue we implement a difference-in-differences (DD) design relying on taxpayers' differential exposure to newly-appointed CAs.

We use trade linkages to construct two distinct groups of firms: those who are connected to newly-appointed CAs—our Treatment group, T—and those who are not—our Control group, C. Firms in T have commercial ties with CAs appointed in November 2016 and therefore experienced an increase in the number of CAs in their network at the time of the reform. Firms in C only trade with pre-existing CAs and did not experience a change in the number of CAs operating in their network of commercial partners. In an ideal setting the two groups would be constructed based on linkages observed at a baseline period. However, we face the constraint that we do not observe trade links before one of the firms becomes CA. This implies that, for firms connected to new CAs, the first observed transaction is in November 2016 at best. To circumvent this issue we assume that there is a 6 month window of time where firms do not drop their partners. Then, we classify into T any firm who has a transaction with a newly-appointed CA between November 2016

and April 2017 while firms in C are those who only have transactions with pre-existing CAs during the same period.

By comparing T vs C we compare a group that experienced a change in the way taxes are collected—from “direct payment” to “withholding”—against a group that did not. Under the assumption that the two groups would have behaved similarly in the absence of the policy change, their comparison will deliver the causal effect of an increase in the share of tax collected through withholding on taxpayer behavior. Coefficient β_3 in the following regression captures this effect

$$y_{it} = \beta_1 \cdot Treat_i + \beta_2 \cdot Post_t + \beta_3 \cdot Treat_i \cdot Post_t + \theta_i + \gamma_t + \varepsilon_{it} \quad (1)$$

where y_{it} is the outcome for firm i in period t , $Treat_i$ is an indicator for firms in the Treatment group, $Post_t$ is an indicator for time periods after November 2016 (the reform date), θ_i are firm fixed-effects, γ_t time period fixed-effects. We focus on the following outcomes. To gauge the first stage of the reform on taxpayers, we look at the number of CAs (extensive margin) and share of tax withheld (intensive margin). The behavioral response is captured by gross income (sales), and we also look at tax liability. Although the data is provided at the monthly level we aggregate it to quarters in order to avoid some issues that are present in the most granular definition, such as: firms reporting zero sales in a given month and large seasonal fluctuations in economic activity related to the combination of the holiday season and the end of the fiscal year. In doing so, we define quarters relative to the reform—i.e., the first quarter corresponds to the months of November 2016 through January 2017. We restrict the time frame of our regressions to all available quarters in the pre-period (starting November 2015) up to the last quarter of 2019 (ending January 2020, before the COVID19 Pandemic began to take a toll on economic activity). Once the time variable is defined in this way, construct a balanced panel of firms who report positive sales at least half of the periods in which they are present in the data.

We complement the DD estimates with the following Event Study analyses

$$y_{it} = \sum_{\tau=-q}^{-2} \delta_\tau \cdot D_{i\tau} + \sum_{\tau=0}^m \beta_\tau \cdot D_{i\tau} + \theta_i + \varepsilon_{it} \quad (2)$$

where $D_{i\tau}$ is an Event Study indicator for each quarter relative to the baseline period (November 2016-January 2017). The coefficients δ_τ and β_τ estimate the effect on the outcome of interest for each period relative to the baseline ($\tau = -1$), net of firm fixed effects. This specification provides compelling visual evidence of the validity of the identification

assumption as the coefficients δ_τ reveal pre-reform trends.

6.1 Empirical Results

Event Study results. We begin by providing evidence for the first-stage of the reform: an increase in the number of commercial partners who are CAs and the related increase in the share of due taxes withheld at source. Figures 9 and 10 plot the event study estimates for the first-stage regressions. In the top panel of each figure we run a separate regression for T and C and plot the coefficients for the time dummies plus the constant. The bottom panel displays the DD coefficients with the corresponding confidence interval assessing the relative difference between the two groups in each month.

A series of features stand out in these plots. First, both cases evolve similarly before the reform. This provides initial evidence in favor of our identifying assumption—absent any policy change, both groups evolved similarly. Second, there is a sharp increase for T at the time of the reform. This suggests that the policy had the intended effect of increasing exposure to CAs throughout their networks of partners, and that the CAs fulfilled their role as tax collectors by withholding at source. Third, while the effect for the number of CAs remains stable through time, the effect on the share of tax withheld decreases to about half of its initial magnitude as we move further away from the time of the event. A potential explanation is that, although firms maintain their commercial partners once they become CAs, they may modify their sourcing decisions to reduce the amount of tax withheld by them.

In Figure 11 we show the analogous plots for the behavioral response to the reform. We plot the coefficients on reported income where we compute its growth relative to $t - 1$, this allows us to re-scale both groups relative to the baseline. The levels plot shows a similar pattern to the first stage figures. Both groups track each other before the reform and there is an increase for T right at the time of the reform. The DD plot allows a clearer appreciation of the immediate increase right at $t = 0$ and further increases over time. Overall, the estimates suggest that treated firms increased their reported gross income by 5 percentage points relative to the control group.

Finally, Figure 12 provides the estimation results for tax liability. The design of the TT implies that tax liability is calculated by applying the corresponding tax rate to the tax base, no deductions are contemplated in the tax. As tax liability is a function of tax base, it would be surprising to find different results in this case. Reassuringly, tax liability displays the same pattern. Taken together, results in Figure 11 and Figure 12 indicate

that the tax administration saw an increase in total tax collected after the reform. This coincides with the macro evidence shown in Section 2.

Difference-in-Differences Results. Table 3 Panel A reports the results from estimating equation 1. We show the first stage and reduced form point estimates for the share of tax withheld and gross income growth relative to the baseline period. This exercise summarizes previous figures and shows that reported sales for treated firms increases by 5 percentage points relative to the control group after the reform is in place. Moreover, in terms of the pass-through, our estimates show that, ...

Interpretation of the results. A potential interpretation of this result is that withheld taxes act as a lower bound on self-reported income: if firms were under-reporting their sales before the reform, having a larger part of the tax remitted by a CA might have forced them to increase the amount they must report.

7 How do firms respond to a decrease in withholding?

We analyze another policy change that took place during the period of our study that—differently to the reform described so far—reduced exposure to withholding for a group of firms. The aim of this exercise is twofold. First, it addresses potential endogeneity concerns that could arise from the previous analysis. Second, it provides an alternative identification channel to the response of firms to withholding.

In September 2018 the City’s government decided to waive withholding by banks to firms whose sales in 2017 were below AR\$ 10 million. This policy was part of a stimulus plan targeted to small and mid-sized enterprises in the context of an economic crisis in the country. The policy measure was initially meant to last six months, although it was later extended for another six months. In a nutshell, the policy: (i) targeted a specific group of taxpayers relying on a pre-determined size threshold; (ii) took place almost two years after the 2016 reform, and (iii) was the “inverse” of the reform analyzed thus far, in the sense that in this case it *reduced* exposure to withholding.

Figure 13 plots the assignment rule for the policy. We take bins of firm sales in 2017 (the policy’s running variable) and calculate the average share of tax withheld by banks in each bin. The share was very similar on either side of the threshold for the periods prior to the policy change and, exactly in September 2018, the share drops for firms below the threshold. The discontinuity is fuzzy because the tax administration denied the benefit to firms flagged as non-compliant (e.g., firms who had failed an audit in the past or

maintained high levels of debt with the administration); these types of measures are more likely to be executed on larger firms.

Conditional on a firm's sales in 2017, being granted the exemption of bank withholding is as good as randomly assigned for those who are close to the cutoff. Therefore, this setting allows us to address any potential endogeneity concerns that remained from the previous analysis.

Moving on to the identification of firms' responses, this setting provides a way to test whether a reduction in the coverage of withholding leads to a fall in reported tax liability. Recall from Section 4 that non-compliant firms will try to underreport up to the point where the self reported sales offset any withheld funds. If the latter fall, then the firm will be able to underreport more. To estimate the effect of lowering withholding exposure on sales, we a similar DD research design as the one summarized by equation 1. In this case, we leverage on the discontinuity created by the assignment criterion and define treatment and control groups based on the location of firms on either side of the cutoff. Firms in the treatment group are those whose sales in 2017 were between AR\$ 5 and 10 million. Firms in the control group are those whose sales in 2017 were between AR\$ 10 and 20 million.

Figure 14 shows the first stage results on the share of taxes withheld by banks. Withholding by banks falls sharply right at the time of implementation for T relative to C. Figure 15 shows the response on sales. Firms who experienced a reduction in withholding reduce their reported sales. We interpret these findings as follows: the policy lead to a reduction in the coverage of third party information for firms in T, whose banks were no longer collecting taxes from them and remitting the funds to the tax administration. This gave firms more freedom to misreport their sales, allowing them to reduce their tax liability.

Higlight the fact that the change was initially temporary. Then it is striking that firms reacted so fast. If they knew withholding was coming back, why they risk revealing that they underreport? There seems to be a behavioral story to it. Perhaps firms use some kind of "mental accounting" and target whatever is withheld. This is sort of surprising given that firms are supposed to be (more) rational relative to individuals. Nevertheless, these are small firms, and they are usually mediated by tax preparers/accountants, who advice them on ways to optimize taxes.

8 Concluding remarks

In this paper, we study the effects of delegating tax collection duties on firms. We analyze the expansion of the withholding scheme used to collect the TT in the City of Buenos Aires, Argentina, which implied an increase in the share of taxpayers' tax liability collected indirectly at source by CAs *in lieu* of direct payments to the tax authority. The setting allows us to estimate the direct effects of expanding the withholding scheme on the firms who are put to the task of collecting taxes, and the indirect effects on those who now face an increase in the amount of taxes being withheld at source.

Our results show that relying on large firms to withhold taxes from their commercial partners is an effective tax collection tool. On one hand, appointing firms to collect taxes did not hurt their activity. On the other hand, substituting direct payments for withholding at source increases reported sales which, in turn, increases tax revenue. Taken together, these findings bolster the argument of recent research by Basri et al. (2019), showing that administrative reforms can be highly effective in raising revenue and building tax capacity.

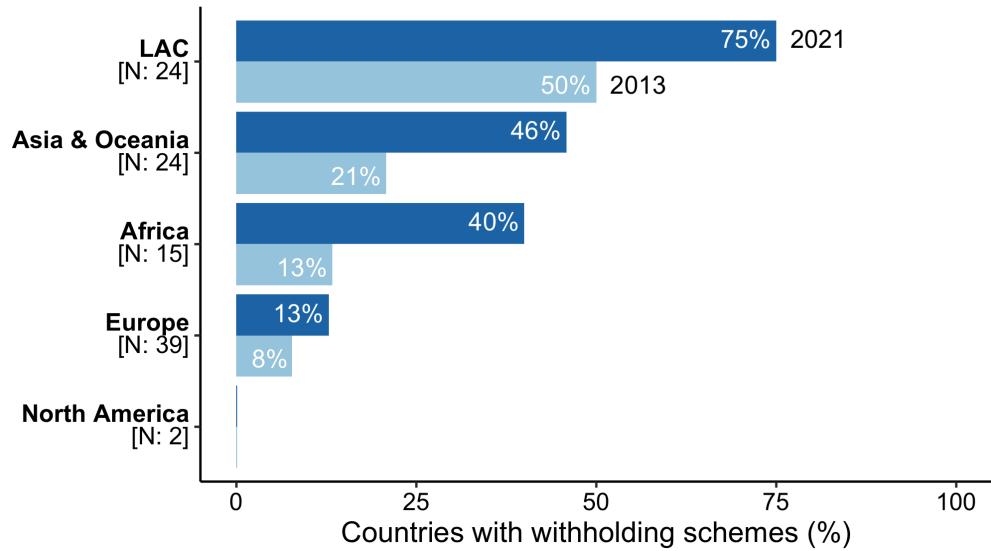
References

- Allingham, M. G. and A. Sandmo (1972). "Income tax evasion: A theoretical analysis". *Taxation: critical perspectives on the world economy* 3, pp. 323–338.
- Almunia, M. and D. Lopez-Rodriguez (2018). "[Under the radar: The effects of monitoring firms on tax compliance](#)". *American Economic Journal: Economic Policy* 10 (1), pp. 1–38.
- Basri, M. C., M. Felix, R. Hanna, and B. A. Olken (2019). "[Tax Administration vs. Tax Rates: Evidence from Corporate Taxation in Indonesia](#)".
- Bergeron, A., G. Tourek, and J. Weigel (2021). "[The State Capacity Ceiling on Tax Rates: Evidence from Randomized Tax Abatements in the DRC](#)".
- Besley, T. and T. Persson (2014). "[Why do developing countries tax so little?](#)" *Journal of economic perspectives* 28 (4), pp. 99–120.
- Best, M. C., A. Brockmeyer, H. J. Kleven, J. Spinnewijn, and M. Waseem (2015). "[Production versus revenue efficiency with limited tax capacity: theory and evidence from Pakistan](#)". *Journal of Political Economy* 123 (6), pp. 1311–1355.
- Brockmeyer, A. and M. Hernandez (2019). "[Taxation, Information and Withholding: Evidence From Costa Rica](#)". *The World Bank*.
- Carrillo, P., D. Pomeranz, and M. Singhal (2017). "[Dodging the taxman: Firm misreporting and limits to tax enforcement](#)". *American Economic Journal: Applied Economics* 9 (2), pp. 144–64.
- Gadenne, L., T. K. Nandi, and R. Rathelot (2020). "[Taxation and Supplier Networks: Evidence from India](#)".
- Gerard, F., J. Naritomi, A. Seibold, and B. Zulian (2019). "[Two-Tier Tax Systems and Firms: Evidence from Brazil](#)".
- Hansen, B., K. S. Miller, and C. Weber (2021). "[Vertical integration and production inefficiency in the presence of a gross receipts tax](#)".
- Keen, M. and J. Slemrod (2017). "[Optimal tax administration](#)". *Journal of Public Economics* 152, pp. 133–142.
- Kleven, H. J., M. B. Knudsen, C. T. Kreiner, S. Pedersen, and E. Saez (2011). "[Unwilling or unable to cheat? Evidence from a tax audit experiment in Denmark](#)". *Econometrica* 79 (3), pp. 651–692.
- Kleven, H. J., C. T. Kreiner, and E. Saez (2016). "[Why can modern governments tax so much? An agency model of firms as fiscal intermediaries](#)". *Economica* 83 (330), pp. 219–246.

- Kopczuk, W., J. Marion, E. Muehlegger, and J. Slemrod (2016). “Does tax-collection invariance hold? evasion and the pass-through of state diesel taxes”. *American Economic Journal: Economic Policy* 8 (2), pp. 251–86.
- Kopczuk, W. and J. Slemrod (2006). “Putting firms into optimal tax theory”. *American Economic Review* 96 (2), pp. 130–134.
- Musgrave, R. A. (1969). “Cost-benefit analysis and the theory of public finance”. *Journal of Economic Literature* 7 (3), pp. 797–806.
- Naritomi, J. (2019). “Consumers as tax auditors”. *American Economic Review* 109 (9), pp. 3031–72.
- Pessina, L. (2020). “Who writes the check to the government does matter: Evidence from firm-to-firm links”. *Job Market Paper*.
- Pomeranz, D. (2015). “No taxation without information: Deterrence and self-enforcement in the value added tax”. *American Economic Review* 105 (8), pp. 2539–69.
- Risch, M. (2021). “Does Taxing Business Owners Affect Employees? Evidence from a Change in the Top Marginal Tax Rate”.
- Slemrod, J. (2008). “Does it matter who writes the check to the government? The economics of tax remittance”. *National Tax Journal*, pp. 251–275.
- Slemrod, J. (2019). “Tax compliance and enforcement”. *Journal of Economic Literature* 57 (4), pp. 904–54.
- Slemrod, J. and S. Yitzhaki (2002). “Tax avoidance, evasion, and administration”. *Handbook of public economics*. Vol. 3. Elsevier, pp. 1423–1470.
- Slemrod, J. B. and T. Velayudhan (2018). “Do Firms Remit at Least 85 Percent of Tax Everywhere?” *Journal of Tax Administration* 4 (1).
- Waseem, M. (2020). “The role of withholding in the self-enforcement of a value-added tax: Evidence from pakistan”. *The Review of Economics and Statistics*, pp. 1–44.

Figures

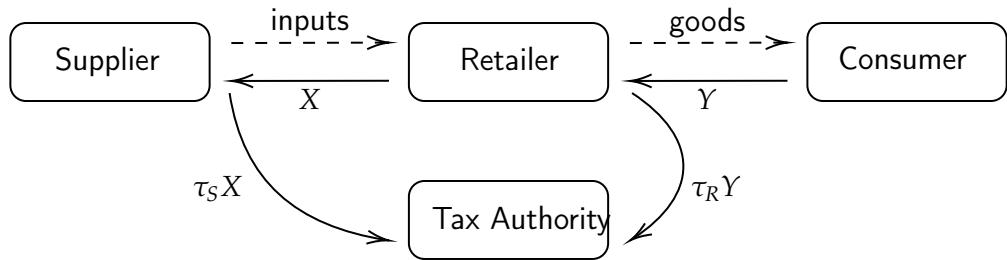
Figure 1: Withholding of indirect taxes by region



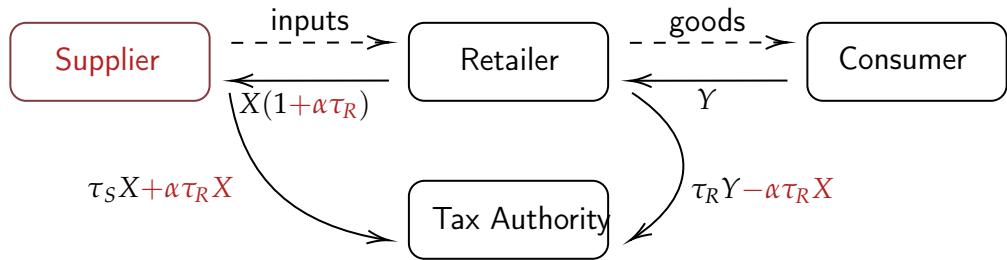
Notes: This figure shows the share of countries in each region with a withholding scheme for indirect taxes in 2013 (light blue bar) and 2021 (blue bar). We construct the figure using text analysis techniques applied to the Ernst & Young's "Worldwide VAT, GST and Sales Tax Guide". For more details, see Appendix B.

Figure 2: Tax collection mechanisms

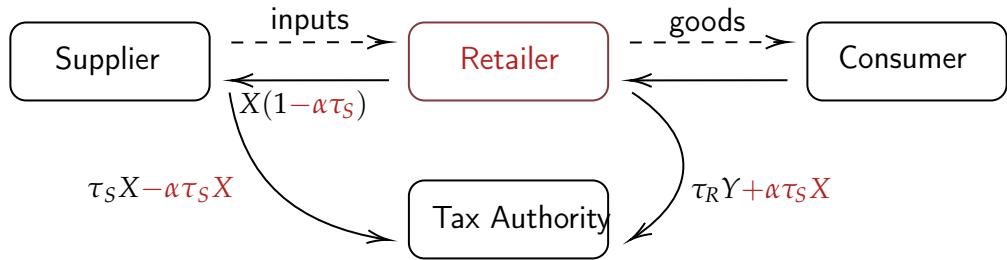
Direct Payment



Withholding by seller (Supplier)

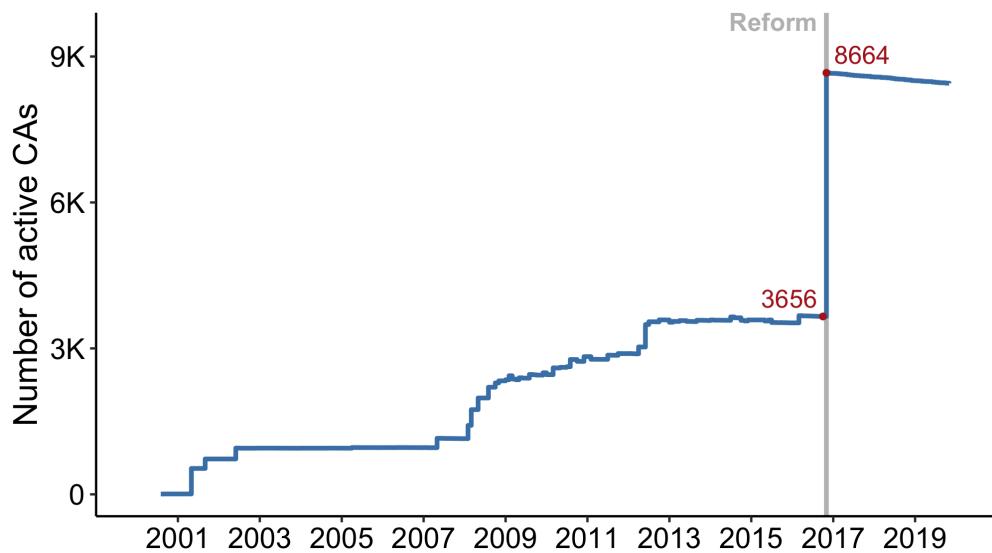


Withholding by buyer (Retailer)



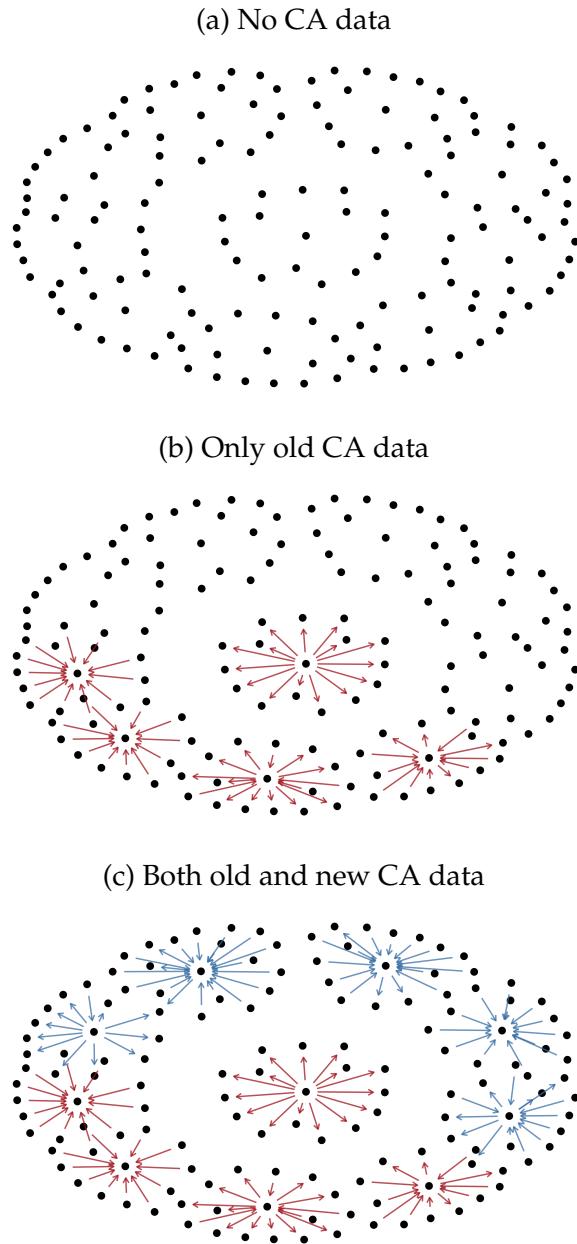
Notes: The diagram summarizes the main features of the “direct payment” and “withholding” tax collection mechanisms in a simplified setting with four agents. Highlighted in red are the agents that get appointed as CA under each scenario, as well as the tax withheld by them.

Figure 3: Number of firms acting as Collection Agents



Notes: The figure shows the number of firms who are part of the Collection Agent register at any point in time. It was constructed using the monthly CA register published by the tax administration as public information. By comparing each successive month, we obtain the actual entry or exit status for each firm, independent of whether they are active in our administrative tax records. See Figure C.3 for more details on how we scrapped the information. “Reform” corresponds to November 2016, the date in which the reform was enacted. Highlighted in red are the number of CAs in October 2016 and November 2016.

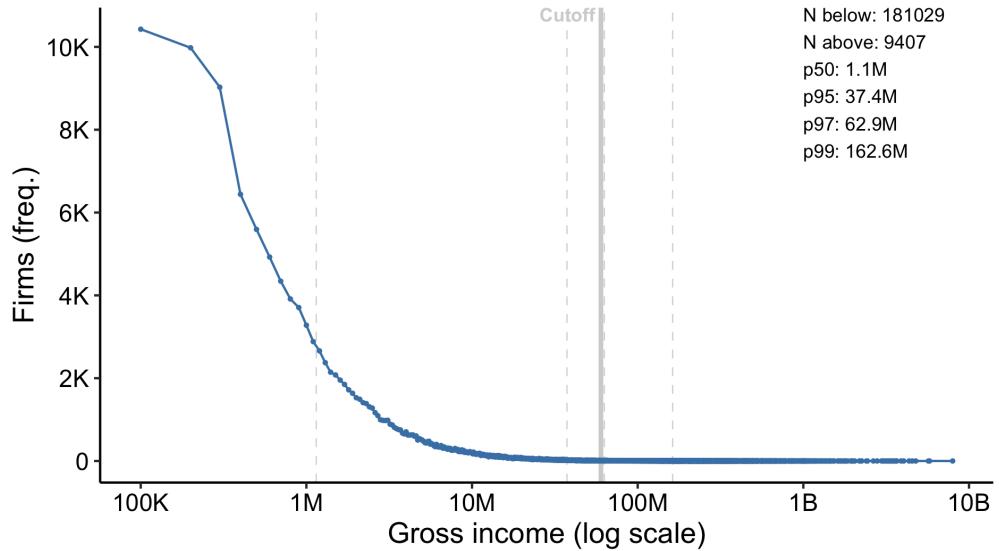
Figure 4: Networks



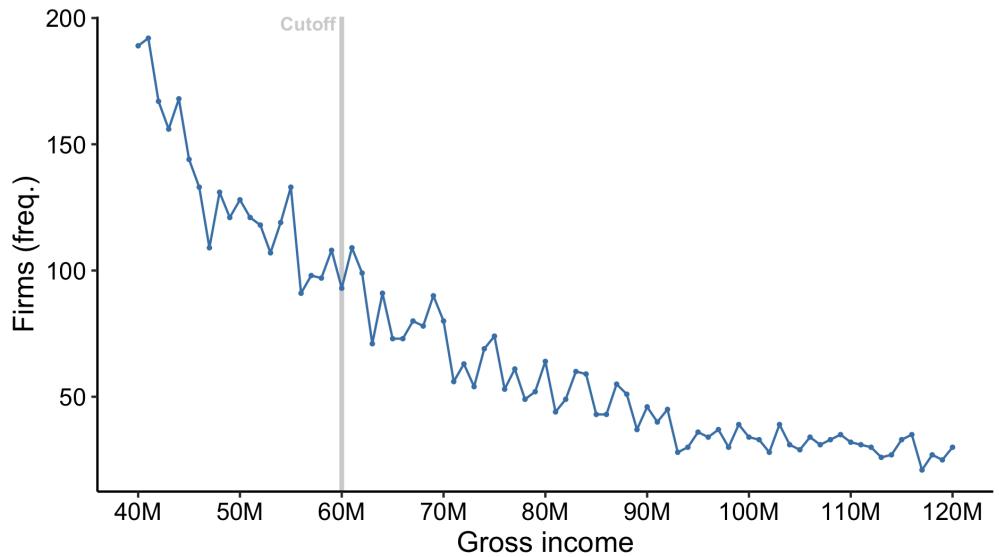
Notes: These figures illustrate the construction of the business to business linkages in our data. Each node represents a real firm, and the connecting edges represent real linkages. The central nodes, from which the edges irradiate, correspond to firms acting as CA while the direction of the edges signify whether the CA is a seller (outgoing) or a buyer (incoming) in the transaction. Notice that the same CA can simultaneously withhold sales or purchases and thus have arrows pointing in both directions.

Figure 5

(a) Pre-reform gross income distribution

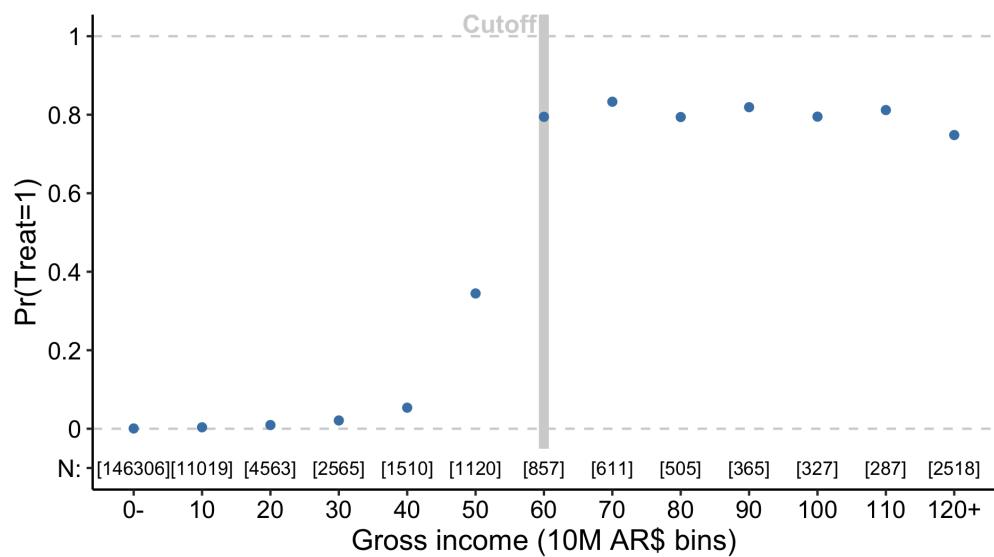


(b) Zooming in around appointment rule cutoff



Notes: Gross income bins of size 100K and 1M AR\$ ($\sim 8K$ and $\sim 80K$ \$), respectively; “Cutoff” indicates the location of income threshold; Dashed lines indicate p50, p95, p97, p99, respectively; Showing relevant part of support in each plot.

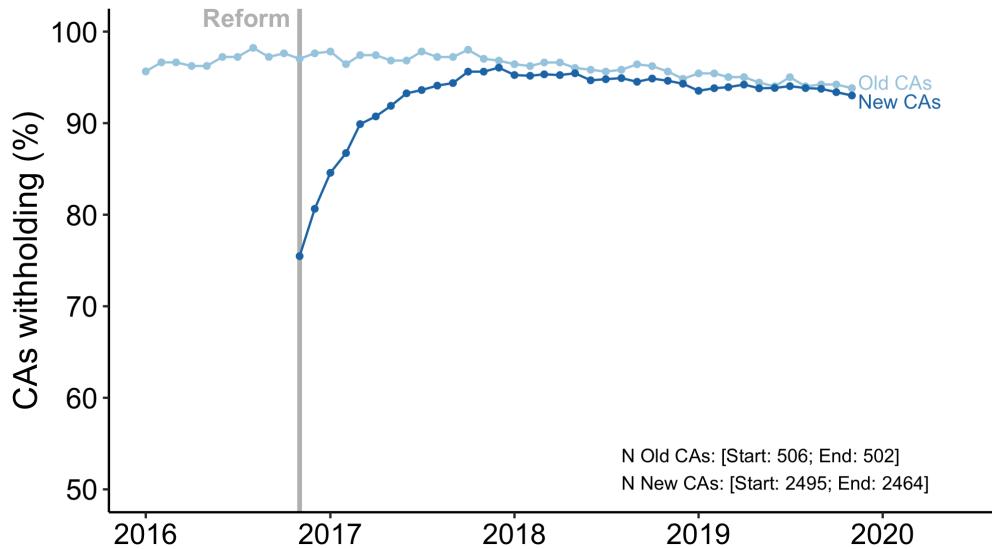
Figure 6: Probability of appointment as Collection Agent



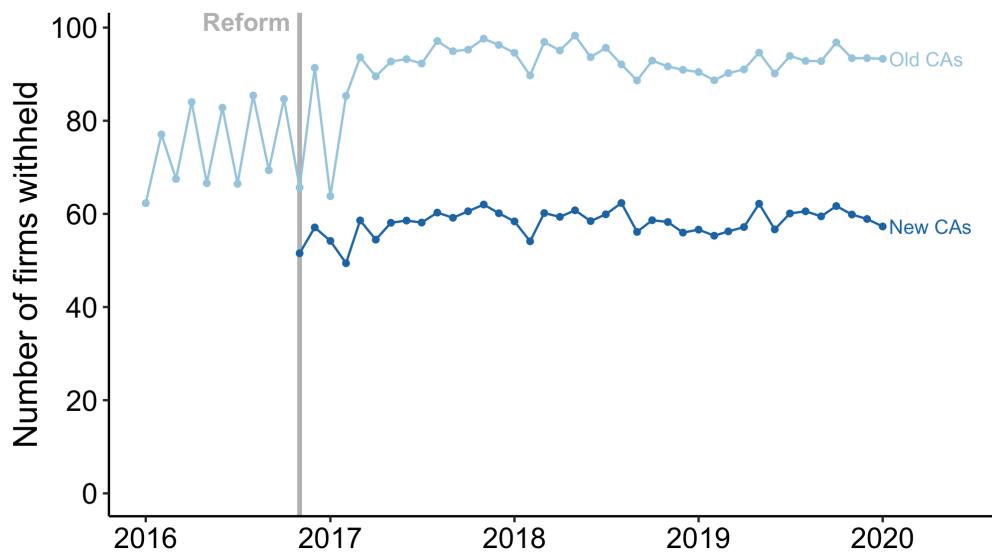
Notes: The figure shows the probability of appointment as Collection Agent by gross income bin. The reform enacted in November 2016 stated that firms whose gross income in 2015 was greater than 60M AR\$ would be automatically appointed. “Cutoff” indicates the location of this threshold. The numbers in brackets indicate the total number of firms in each bin.

Figure 7: Withholding by CAs: extensive and intensive margins

(a) Extensive margin: Proportion of CAs withholding from clients

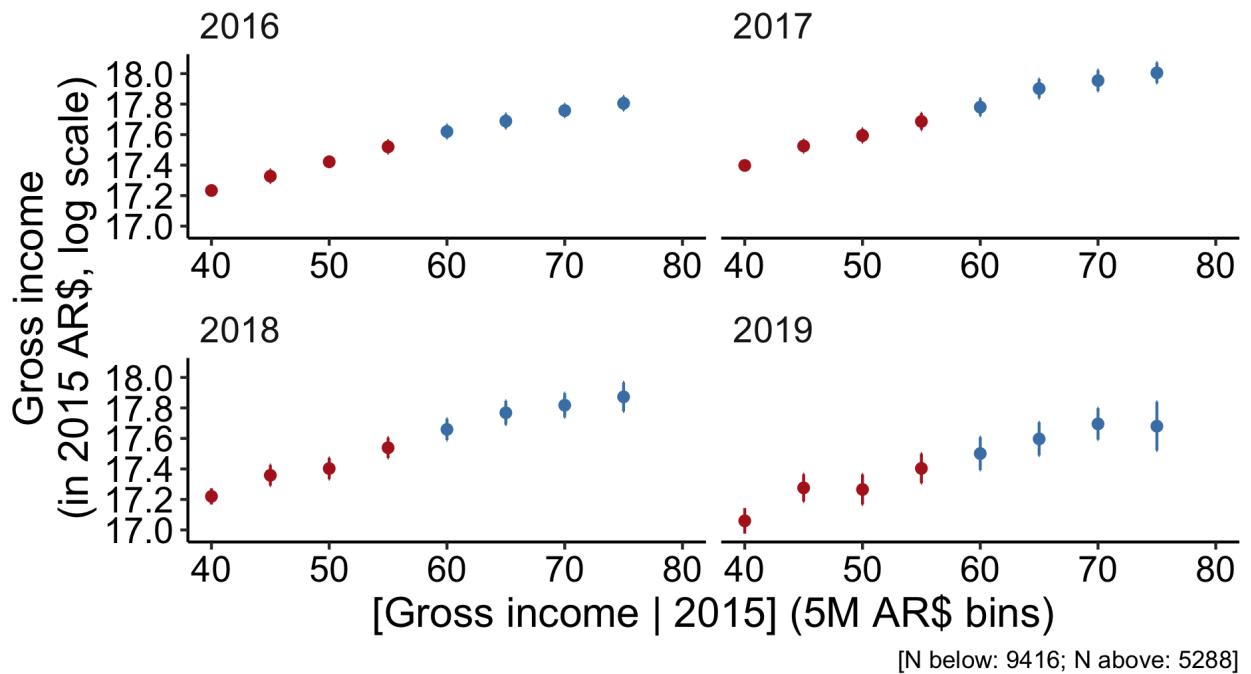


(b) Intensive margin: Average number of firms withheld by CAs



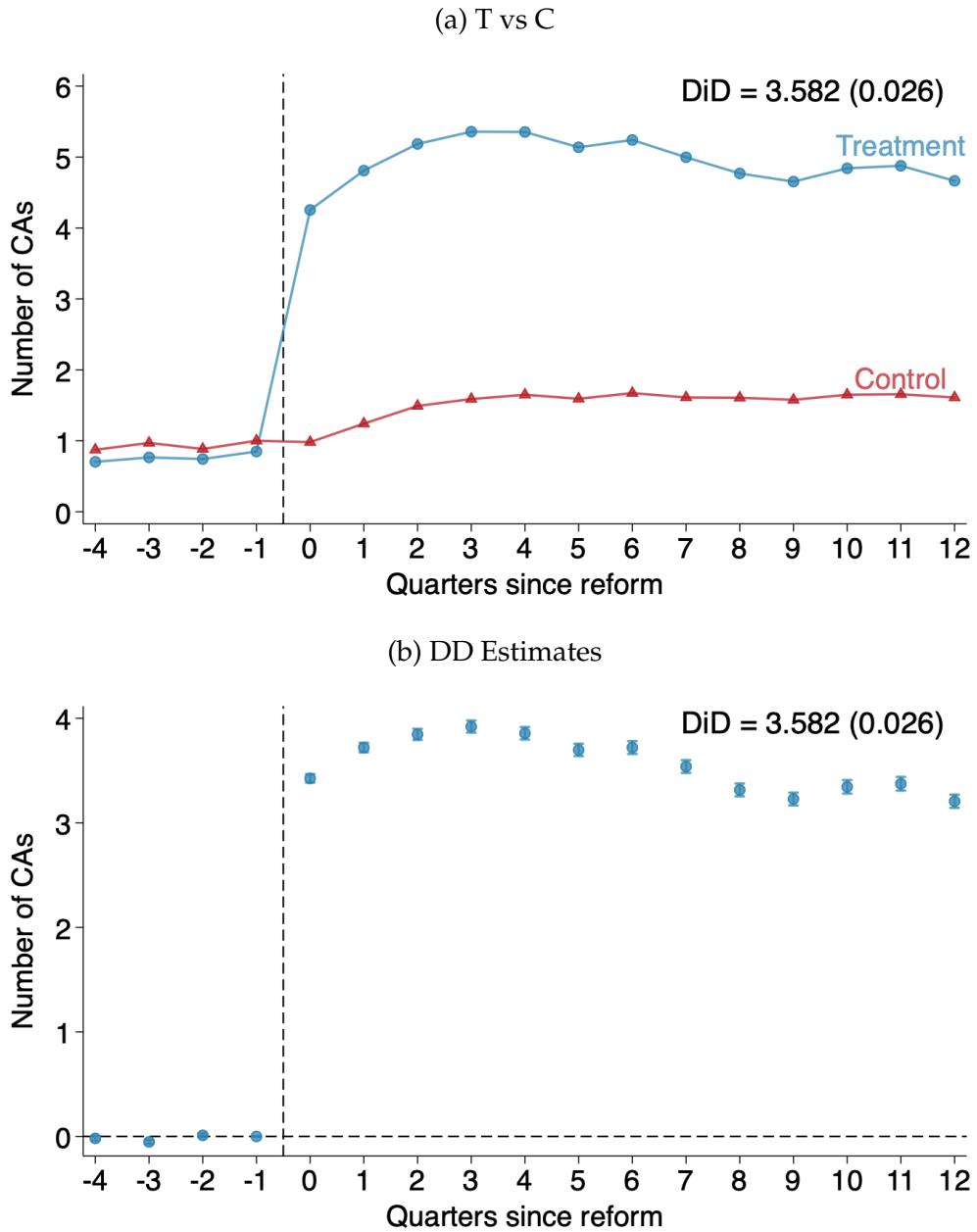
Notes: xx

Figure 8: Gross income of newly appointed CAs vs non-eligible firms



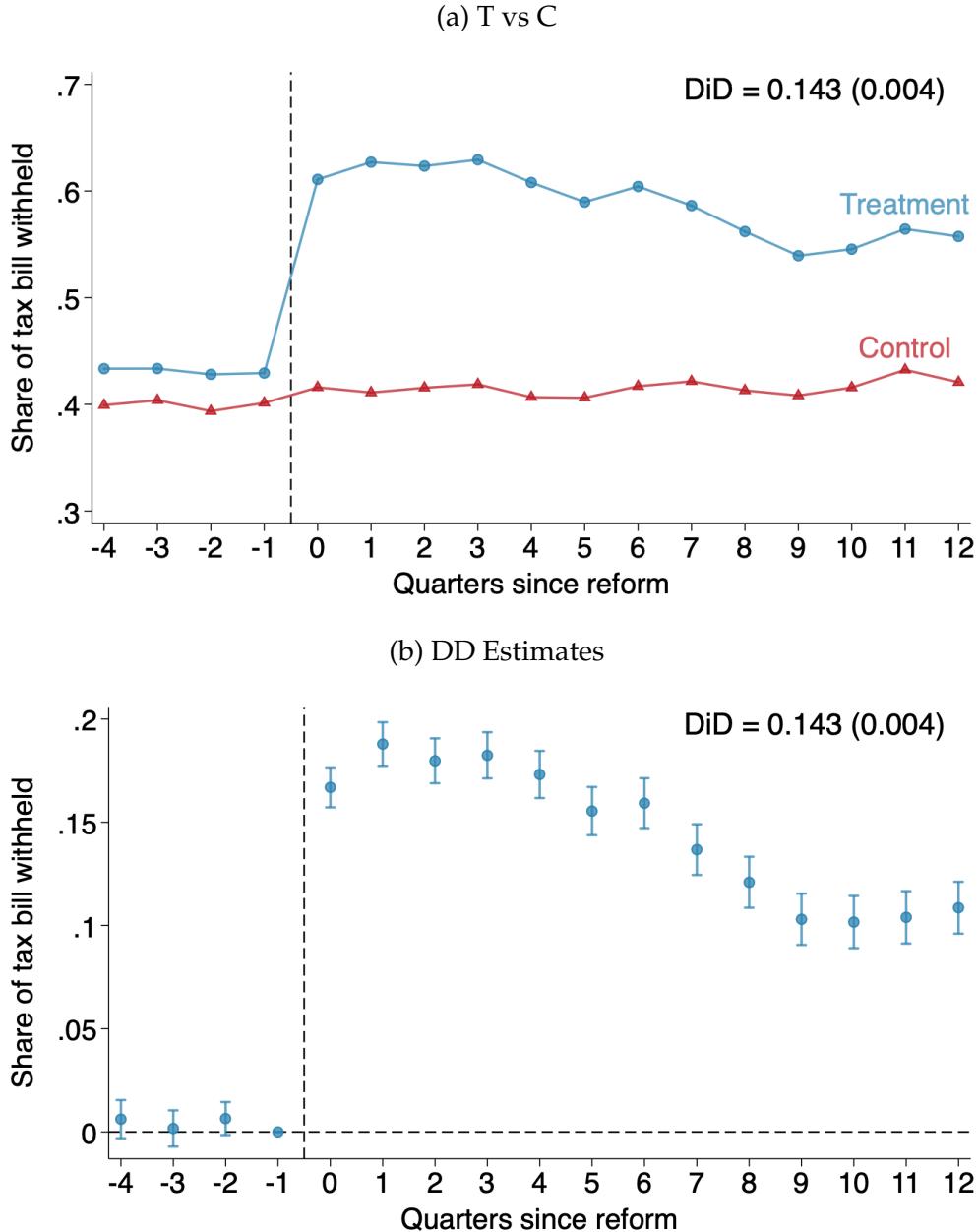
Notes: This graph plots the log annual gross income in constant pesos of 2015 for 8 equally-spaced bins (width AR\$ 5 millions) of the running variable: the annual gross income in 2015. The top left panel corresponds to the year of 2016 excluding the two treated months, November and December. Hence, it can be interpreted as a placebo test because it only contains pre-reform sales. The top right panel and bottom panels show annual sales in 2017 through 2019. The sample is composed of a balanced panel of firms. We exclude outliers (growth rate below percentile 1 and above percentile 99) from our calculations.

Figure 9: Number of CAs



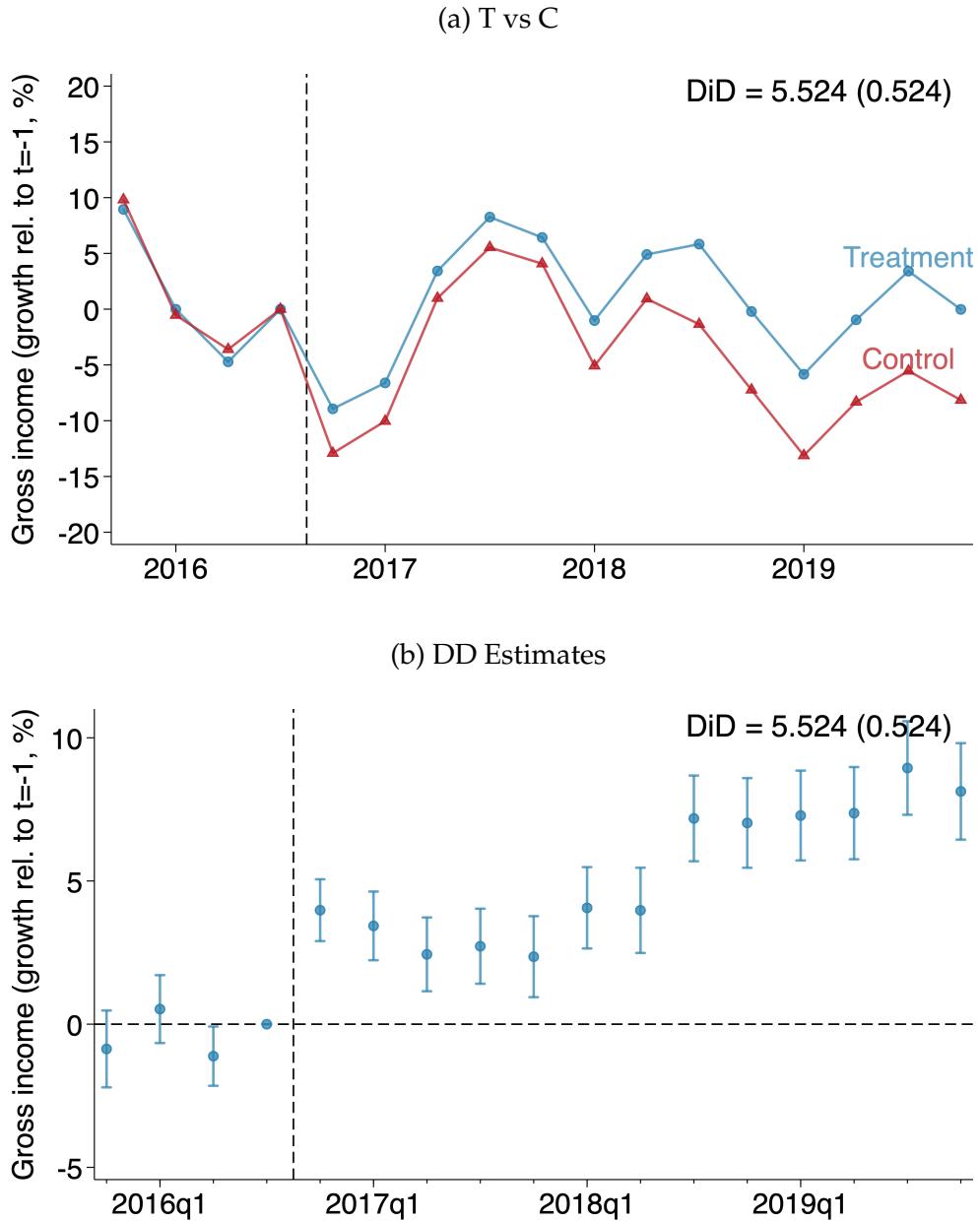
Notes: The LHS panel corresponds to the regression of the outcome of interest on the full set of firm and month fixed effects, dropping the dummy for the baseline period. The figure plots the coefficients on the constant and the time dummies of these regressions, run separately for the two groups of firms. Each coefficient represents the average within-firm number of CAs relative to the baseline month. The RHS panel is the DD analogue of the LHS panel. All regression specifications contain a balanced panel of firms who filed taxes at least six months every year.

Figure 10: Tax withheld



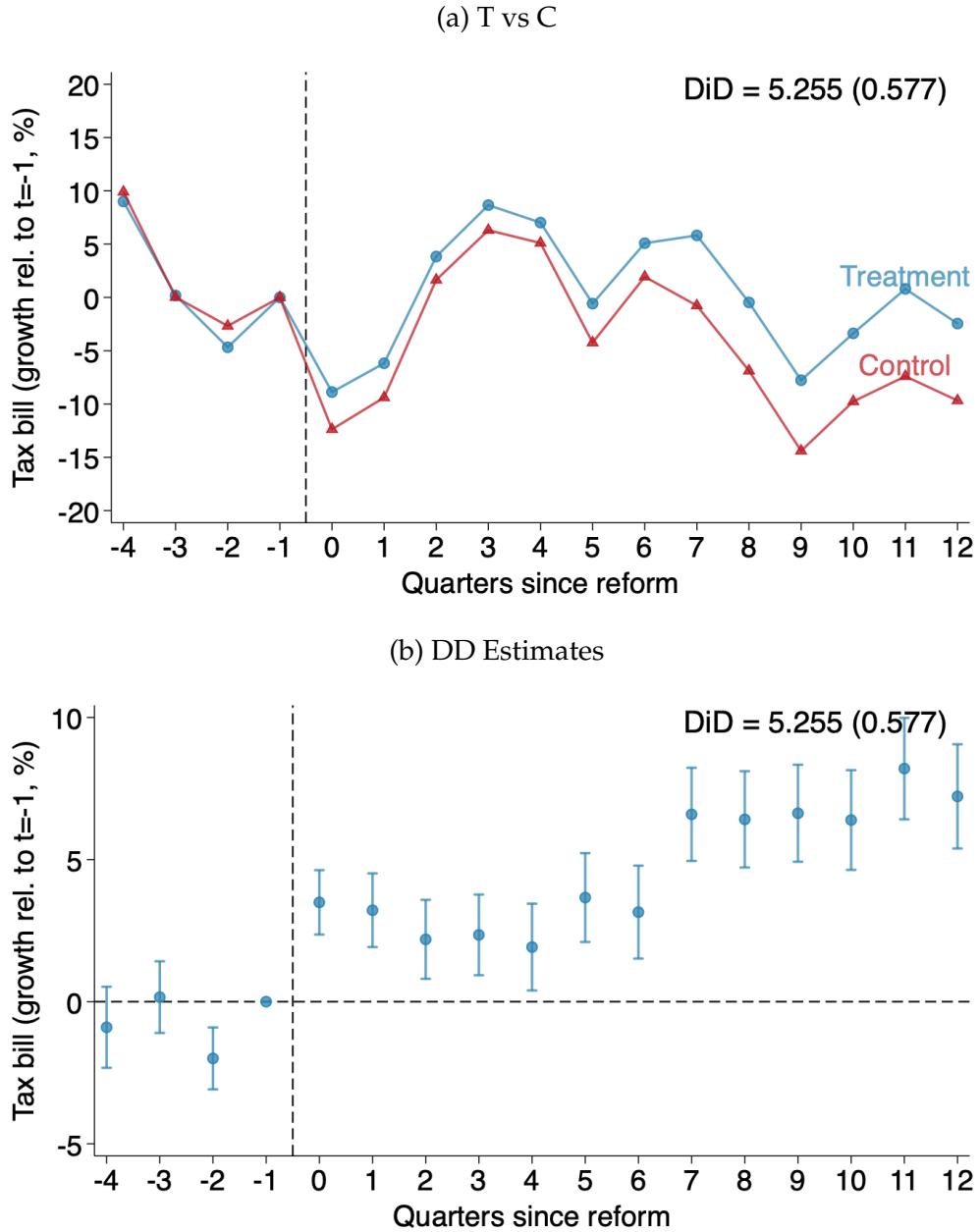
Notes: The LHS panel corresponds to the regression of the outcome of interest on the full set of firm and month fixed effects, dropping the dummy for the baseline period. The figure plots the coefficients on the constant and the time dummies of these regressions, run separately for the two groups of firms. Each coefficient represents the average within-firm change in tax withheld as a share of total tax liability relative to the baseline month. The RHS panel is the DD analogue of the LHS panel. All regression specifications contain a balanced panel of firms who filed taxes at least six months every year. Outcomes are winsorized at percentiles 5 and 95.

Figure 11: Gross income growth



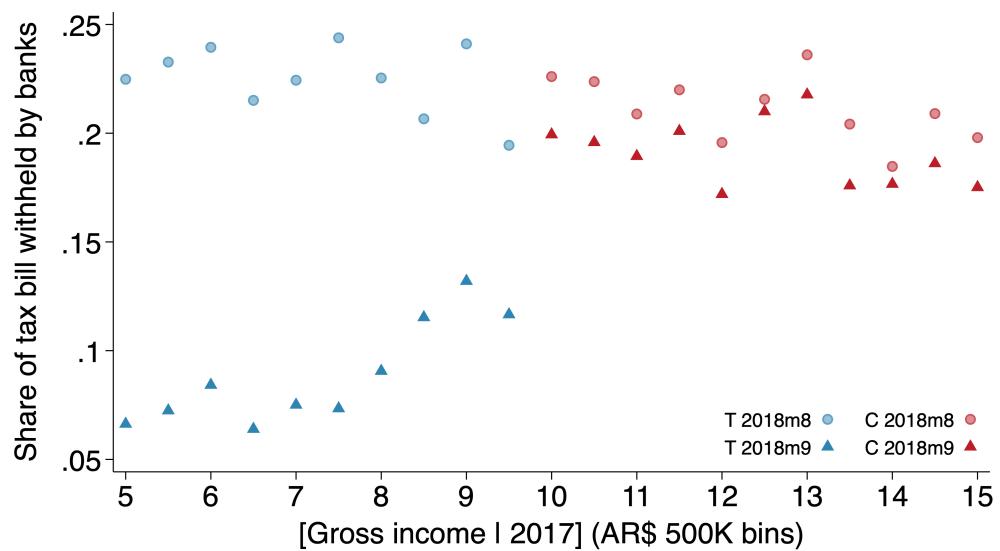
Notes: The LHS panel corresponds to the regression of the outcome of interest on the full set of firm and month fixed effects, dropping the dummy for the baseline period. The figure plots the coefficients on the constant and the time dummies of these regressions, run separately for the two groups of firms. Each coefficient represents the average within-firm gross income growth relative to the baseline month for the corresponding group of firms. The RHS panel is the DD analogue of the LHS panel. All regression specifications contain a balanced panel of firms who filed taxes at least six months every year.

Figure 12: Tax liability growth



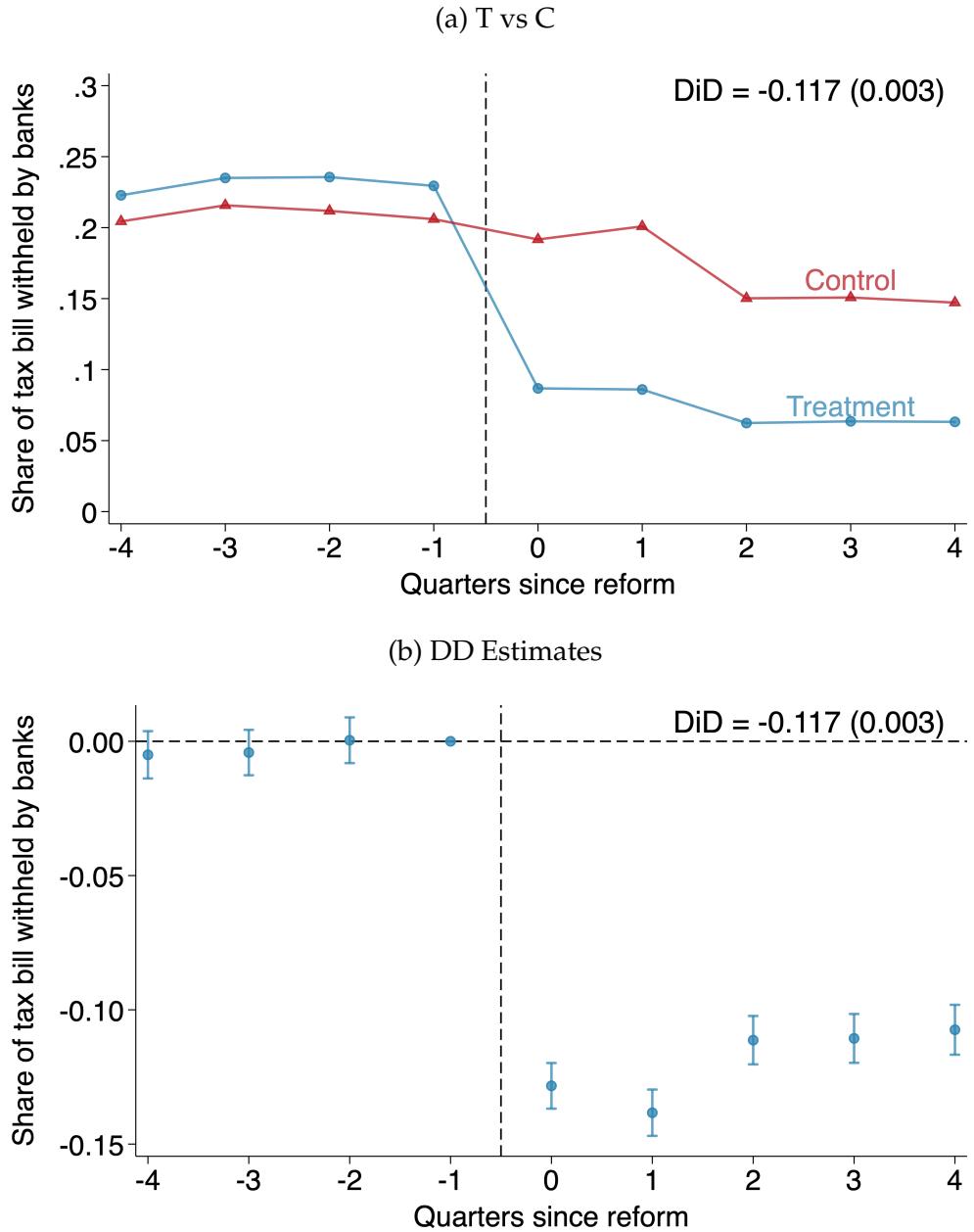
Notes: The LHS panel corresponds to the regression of the outcome of interest on the full set of firm and month fixed effects, dropping the dummy for the baseline period. The figure plots the coefficients on the constant and the time dummies of these regressions, run separately for the two groups of firms. Each coefficient represents the average within-firm tax liability growth relative to the baseline month for the corresponding group of firms. The RHS panel is the DD analogue of the LHS panel. All regression specifications contain a balanced panel of firms who filed taxes at least six months every year.

Figure 13: Tax withheld by banks



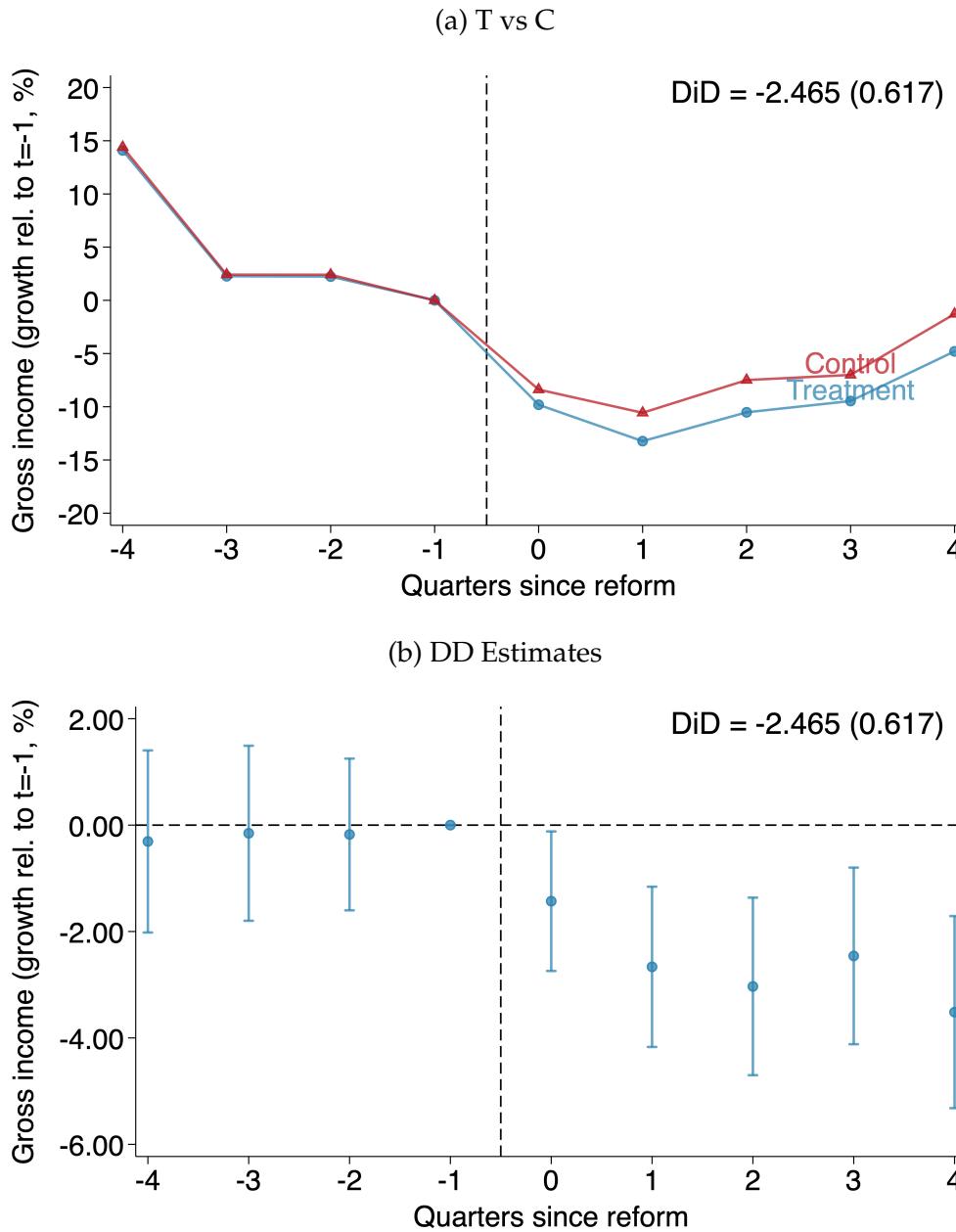
Notes: xx

Figure 14: Tax withheld by banks



Notes: The LHS panel corresponds to the regression of the outcome of interest on the full set of firm and month fixed effects, dropping the dummy for the baseline period. The figure plots the coefficients on the constant and the time dummies of these regressions, run separately for the two groups of firms. Each coefficient represents the average within-firm change in tax withheld as a share of total tax liability relative to the baseline month. The RHS panel is the DD analogue of the LHS panel. All regression specifications contain a balanced panel of firms who filed taxes at least six months every year. Outcomes are winsorized at percentiles 5 and 95.

Figure 15: Gross income growth



Notes: The LHS panel corresponds to the regression of the outcome of interest on the full set of firm and month fixed effects, dropping the dummy for the baseline period. The figure plots the coefficients on the constant and the time dummies of these regressions, run separately for the two groups of firms. Each coefficient represents the average within-firm gross income growth relative to the baseline month for the corresponding group of firms. The RHS panel is the DD analogue of the LHS panel. All regression specifications contain a balanced panel of firms who filed taxes at least six months every year.

Tables

Table 1: Summary statistics

	No. Firms	Gross Revenue				Tax Liability			
		p5	p50	mean	p95	p5	p50	mean	p95
Panel A: Full sample									
Firms	183,503	0	42	249	1,052	0	1	6	28
Collection Agents	9,366	3	2,757	19,209	39,947	0	85	942	1,451
Panel B: Estimating sample									
Firms	80,208	7	123	415	1,664	0	3	11	45

Notes: This table reports summary statistics calculated in the period January–October 2016, before the expansion of the withholding regime. Gross revenue and Tax liability are expressed in thousand of Argentinian pesos from 2016. Panel A corresponds to the full sample. Panel B is the estimating sample used for the taxpayer responses in Sections 6 and 7.

Table 2: Regression discontinuity estimations

	First stage		Reported sales			
	(1)	(2)	2016	2017	2018	2019
RD estimate	0.323*** (0.054)		0.845 (0.588)	0.555 (0.621)	0.839 (0.571)	0.554 (0.559)
Number of firms	4,098		3,235	3,195	3,151	3,102

Notes: This table reports the first stage and the 2SLS estimates from a fuzzy RD specification. The running variable corresponds to the annual gross income in 2015. The reform to the withholding system enacted in November 2016 stated that firms whose gross income in 2015 was greater than AR\$ 60M (with some exceptions) would be automatically appointed as collection agents (CA). Column 1 shows the first-stage change in the probability of a firm above the threshold to be appointed as a CA in November 2016. This first stage was graphically shown in Figure 6. Columns 2 through 5 correspond to the logarithm of reported annual sales. Column 2 excludes the two treated month of 2016, November and December. Therefore, the results of this column can be interpreted as a placebo test as it only contains pre-reform sales. *, **, *** denote statistical significance at the 10, 5 and 1 percent level. The regression is estimated for the range [40, 60) to the left, and [60, 80) to the right of the cutoff using a linear approximation and a triangular kernel. RD estimates are computed with the `rdrobust` routine from **Calonico-etal-2017**.

Table 3: Difference-in-differences estimations

	First stage (1)	Reported sales (2)	Tax revenue (3)
Panel A: Increase in withholding			
<i>Treat · Post</i>	0.143*** (0.004)	5.524*** (0.524)	5.255*** (0.577)
Observations	1,221,000	1,221,000	1,221,000
Number of firms	75,710	75,710	75,710
\bar{y}_{t-1}	11.623	965.219	33.437
Panel B: Decrease in withholding			
<i>Treat · Post</i>	-0.108*** (0.003)	-2.798*** (0.701)	-4.285*** (0.776)
Observations	142,913	142,913	142,913
Number of firms	16,006	16,006	16,006
\bar{y}_{t-1}	7.236	1156.487	33.105

Notes: This table reports the first stage change in withholding and the reduced-form estimates from a difference-in-differences specification. Column 1 shows the first stage results. For Panel A, this corresponds to the share of taxes withheld by CAs, while for Panel B it is the share of taxes withheld by banks. Columns 2 and 3 correspond to the gross income and the tax revenue growth relative to the baseline period (thirds quarter of 2016 and third quarter of 2018, respectively). The last row of each panel shows the average value of the variable reported in columns at baseline. Standard errors clustered at the firm level are reported in parentheses. *, **, *** denote statistical significance at the 10, 5 and 1 percent level.

Appendices

A Turnover Tax: further details

The Turnover Tax (TT)—aka Gross Receipts Tax—is an indirect provincial tax imposed in each of the 24 jurisdictions in Argentina (including 23 provinces and the City of Buenos Aires). It is levied on the gross sale of goods or services with no possibilities to deduct costs. It therefore applies to all the transactions taking place in the supply chain—i.e., business to business and business to consumer transactions. This type of tax is simple to collect but creates “cascading effects” whereby final goods are taxed multiple times throughout production.¹ The TT represents the main source of own revenue in all the jurisdictions (about 75% of total tax receipts in the City of Buenos Aires).² Each jurisdiction issues their own sub-national tax laws and TT regulations in relation to the tax base, tax rates, apportionment rules, collection schemes, exemptions and promotional regimes applicable to each activity.

Tax rates. The tax rates typically vary from 3.5% to 5% for the sale of goods and services depending on the taxpayer activity, annual turnover from the previous year, and the place where the transaction takes place (inside vs outside the province). For example, in 2016, retailers with 2015 revenue less than 49 million pesos faced a 3% tax rate on total sales. This rate increased to 4% if their 2015 revenue was greater than 49 million pesos (i.e., it operates as a notch rather than a kink). Importantly, firms selling out-of-province might be taxed at higher tax rates (as if there were internal customs). This type of firms are informally called “Foreign Jurisdiction” firms. When filing taxes, multi-activity firms must discriminate their tax base by the type of activities they carry out and apply the corresponding tax rate in each case. For example, a large footwear manufacturer that also sells shoes to final consumers could be taxed at 1% for their manufacturing sales and 4% for their retail sales.

Local vs multi-province firms. Taxpayers are classified as *local* when they only sell goods or services within the jurisdiction (e.g., a corner shop) or *multi-province* when they operate both inside and outside the jurisdiction (e.g., a manufacturing from another province selling goods in the City of Buenos Aires). This set of firms is part of the so-called Multi-

¹In environments with limited enforcement capacity this type of tax provides limited opportunities for firms to engage in tax evasion relative to profit-based taxes like the VAT and corporate income tax (Best et al., 2015).

²The other taxes levied at the provincial level are the stamp tax, property tax, car tax, and other minor duties.

lateral Agreement.

Apportionment formula. For firms operating in multiple provinces, there exists an apportionment system to distribute the tax base across places. The formula is based on a sales factor and an expenditure factor determined at year t-1 (unified coefficient). In particular, 50% of total sales is distributed according to the ratio of a taxpayer's sales in the taxing province to its overall sales; and the other 50% is distributed according to the ratio of a taxpayer's expenditures in the taxing province to its overall expenditure.

How is the TT collected? The TT can be paid directly by taxpayers (standard method) or collected indirectly by third-parties (withholding method). In the standard method, taxpayers log on to the tax administration's website at the end of the month, self-report the gross revenue accrued in each activity over that month, the system applies the corresponding tax rate and generates a payment coupon. Under the withholding method, the Tax Administration can designate firms, banks, or credit and debit card companies to operate as collection agents (CAs). In the case of firm CAs, every time a taxpayer purchases inputs from CAs or sells goods or services to CAs, part of the invoice amount is withheld by the CA and remitted to the tax authority. In the case of banks, every time a payment is deposited in the bank account, the bank withholds part of it and remits the funds to the tax authority. In the case of card CAs, every time a transaction is made with debit or credit cards, the card company withholds part of it. In all these cases, taxpayers still have to file taxes at the end of each month, but the withheld amount constitutes a credit in favor of the taxpayer who can discount it from the tax liability before making any payments.

Withholding and reverse withholding. Withholding can operate both downstream and upstream. Upstream withholding is similar to that used for the collection of personal income tax in the US. In this case the payer in the transaction (retaining agent) withholds a fraction of the total amount of the sale and remits it to the Tax Administration in concept of future tax payments from the payee. In the downstream case, the supplier (perceiving agent) withholds from the payer by adding the withheld tax to the total sale and remits it to the Tax Administration in concept of future tax payments from the payer. The remitted funds are accumulated in the withheld taxpayer's account and can be used to deduct from future outstanding tax liabilities. Importantly, in all these cases, the CA must report very detailed information to the tax authority: the firms they traded with, the tax identifier, the total amount of the transaction, and the withheld amount. This information is automatically available for taxpayers when they file their monthly tax returns using the official processing software. If a CA fails to report a transaction, then taxpayers can report it themselves so that they can claim the advance payment back. When that happens, CAs

can face severe penalties. Hence, CAs have strong incentives to perform their collection duties accurately.

Over-withholding. In practice, the sum of withheld amounts could be smaller or greater than the tax liability. On the one hand, not all the transactions are subject to withholding and, generally, the withholding rate is lower than the statutory tax rate. This implies that taxpayers are usually able to compensate a fraction of the tax liability with the advance payments withheld at source. However, it could happen that the monthly withheld amount exceeds the monthly liability, thus generating a credit balance in favor of the taxpayer that can be rolled over for future tax payments. These funds cannot easily be withdrawn and are essentially an interest-free loan from taxpayers to the Tax Authority. The main reason why this could happen is that credit card companies and banks also act as collection agents. So, a given transaction between two trade partners might be subject to withholding multiple times (e.g., by a supplier, by credit card companies, and finally when the money is deposited in a bank account).

TT collection over time. Until the early 2000s, the tax was basically collected via direct monthly payments. From that point, the tax administration started appointing firms to act as collection agents on an ad-hoc basis—usually targeting the biggest firms in the economy. As the tax authority developed their IT systems, the appointment of firms became more widespread and eventually expanded their net in a discrete way as was shown in Figure 3. That is precisely the main reform we exploit in our analysis.

Who becomes a collection agent (CA)? The 2016 reform we analyzed replaced the ad-hoc way of appointing collection agents with a simple revenue-based rule: Firms whose 2015 revenue was greater than 60 million pesos were automatically enrolled as collection agents (Resolution 364/AGIP/2016). These firms could act as both *perceiving agent*, when selling goods or services, and *retaining agent* when purchasing goods or services.

When a firm is enrolled as CA it must start using a separate tax collection software that incorporates withholding to its sales. This software has an updated register of all firms that are liable to be withheld and also determines the rate that should be applied. The withholding rate is applied to the amount before taxes. Firms must transfer the withheld funds every fifteen days along with an annex containing the details of every transaction: tax ID for the withheld party, total transaction amount, and amount perceived or retained. Anecdotally, given that appointed firms are large and formal, they do not find it very complicated to adapt to this new system.

Federal Taxes. Firms are also subject to federal taxes. Depending on the firms' revenue,

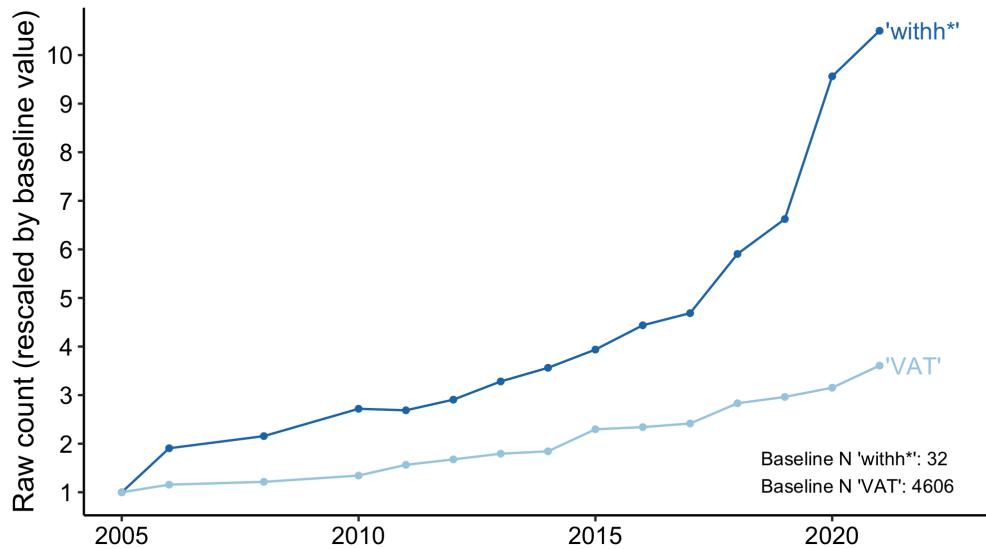
it will have to file a monthly VAT return or be part of a simplified regime. The VAT base is total value added, this implies that firms can deduct purchases from their total revenue. On the other hand, firms in the simplified regime pay a fixed amount depending on thresholds determined by gross income, the same variable used to determine the tax rate and tax liability of the turnover tax.

B Text analysis using Ernst & Young guide for indirect taxes

We explain in more detail the process followed to construct the database used in Figure B.1. We started by downloading Ernst & Young's "Worldwide VAT, GST and Sales Tax Guide". We split each of these guides into country chapters where possible (2013 onwards). Then, we created a text analysis program to scrape information from each chapter. For each country in the guide and every year available, the program searched for terms related to "tax withholding". In particular, we searched strings containing the term "withh*" (i.e., withholding, withheld, withhold, withholdings, withholder, etc.). The program created an binary indicator equal to one if a country's chapter contained any of these terms. As a benchmark, we repeated the text analysis exercise searching for terms related to the value added tax (VAT), since this is also a widespread tax instrument.

In the figure below, we show the raw counts (scaled by the 2005 value) of the total number of matches per guide (i.e., the sum across countries). The dark blue line corresponds to terms related to withholding and the light blue line to VAT. We observe a sharp increase in the number of matches over time. Moreover, the trend accelerates from 2018 onward. This suggests that withholding regimes are become more widespread. Moreover, from Figure B.1, we can see that this is particularly pronounced in low- and middle-income countries who usually present a weaker tax capacity. The results of our paper are encouraging and suggest that this might be a good development of tax systems.

Figure B.1: Number of matches per document:
“withh*” vs “VAT” (used as benchmark)



Notes: This figure summarizes the results from our text analysis exercise. It plots the time series of the number of times that our program detects terms related to withholding (dark blue) or VAT (light blue). Each series is re-scaled by the level in the year 2005. Source: Own elaboration based on [Ernst & Young's "Worldwide VAT, GST and Sales Tax Guide"](#). Our R code is available upon request.

C Additional Figures and Tables

Figure C.2: Relevance of the reform

ámbito.com

Modificaciones en régimen de retenciones en CABA

EDICIÓN IMPRESA 11 Octubre 2016

Los agentes de retención en el régimen de recaudación reformulado serán designados de oficio cuando superen ciertas cifras de ingresos del año calendario anterior sin advertirse un mecanismo de reclamo.



Por [Jorge A. Carmona y Héctor A. Duguine](#)

El régimen recaudatorio vigente en la actualidad se haya dispuesto por Resolución N° 939/AGIP 2013, con las modificaciones introducidas por la Resolución N° 364/AGIP 2016 y la Resolución N° 421/AGIP 2016. Las modificaciones introducidas, crean un típico régimen general de recaudación de retenciones y percepciones basado en gestión de alícuotas por Padrón de Contribuyentes, intentando emular el vigente en la provincia de Buenos Aires administrado por ARBA con retenciones y percepciones cruzadas, perdiendo los Agentes de Recaudación el beneficio de no ser retenidos y percibidos. Hasta la modificación efectuada por la Resolución N° 364/AGIP 2016, en la CABA los agentes de recaudación eran en todos los casos designados taxativamente por el Fisco, a partir de ahora coexistirá ese régimen con el de padrones de alícuotas para el régimen general de retenciones y percepciones. Además de los nominados, deberán actuar en ese carácter los sujetos que desarrollen actividades en la Ciudad Autónoma de Buenos Aires que obtienen en el año calendario inmediato anterior ingresos por un monto los \$60.000.000. A diferencia de la provincia de Buenos Aires en donde el potencial agente debe inscribirse, en la CABA será designado de oficio cuando supere esa cifra. No advertimos un mecanismo exprés de reclamo.

Con ese objeto deberán los ingresos gravados, exentos y no gravados correspondientes a todas jurisdicciones, netos de impuestos. Se considera que desarrollan actividades en la Autónoma de Buenos Aires aquellos sujetos que posean en esta jurisdicción agencias, representaciones, oficinas, locales y todo otro tipo de explotación, edificio, obra, depósito o similar y quienes se valgan para el ejercicio de su actividad en territorio de la Ciudad Autónoma de Buenos Aires, de los comisionistas, corredores, consignatarios o martilleros. Existe una serie de sujetos excluidos de actuar como agentes de recaudación.

Notes: This figure shows an article from Ambito newspaper that discusses the major expansion of the withholding scheme analyzed in this paper. The title says "Changes in the withholding regime in CABA (City of Buenos Aires)" and the text below reads "Withholding agents in the reformulated collection regime will be appointed *ex officio* when they exceed an income threshold from the previous calendar year and have no right to appeal the appointment". Related newspapers also published articles with FAQ clarifying the details of the reform. For example, see [iProfessional](#). Source: [Ambito.com](#)

Figure C.3: Example of the register of withholding agents published monthly by the Tax Administration

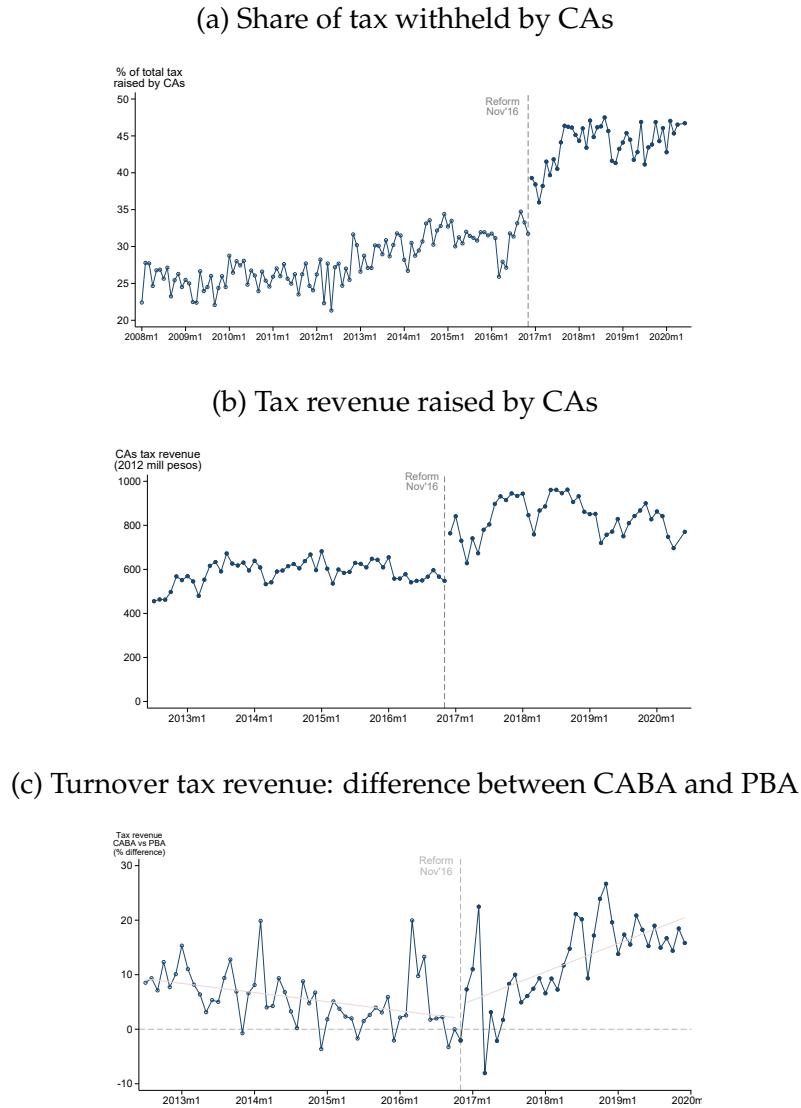


NOMINA AGENTES DE RECAUDACION
MAYO 2017

ORDEN	A.I.	D.V.	CUIT	RAZON SOCIAL	NORMA	VIGENCIA
1248	15404	0	30711230188	CABLETECH ARGENTINA SA	RES. 364-AGIP-2016	01/11/2016
1249	8410	9	30573652084	CABLEVISION SA	RES. 1080-MHGC-07	02/05/2007
1250	9454	2	30500599692	CABOSCH S.A.	RES. 607-AGIP-2009	01/10/2009
1251	15875	0	33504113049	CABOT ARGENTINA SAIC	RES. 364-AGIP-2016	01/11/2016
1252	1543	7	30537745742	CABRALES SOCIEDAD ANONIMA	Res. 1465-SHYF/01	01/09/2001
1253	9506	6	30708969245	CACHAYS A.S.	RES. 50-AGIP-2010	01/03/2010
1254	11537	0	30507400228	CACHI YACO SA	RES. 364-AGIP-2016	01/11/2016
1255	13811	0	30700584131	CACHO SUSPENSION S.A.	RES. 364-AGIP-2016	01/11/2016
1256	15356	0	30710988307	CAICIC SPORTS VISION SRL	RES. 364-AGIP-2016	01/11/2016
1257	9519	5	30544013544	CADENACI SA	RES. 50-AGIP-2010	01/03/2010
1258	9760	8	30578899002	CAFE OYAMBRE SRL	RES. 371-AGIP-2010	01/08/2010
1259	9761	1	30707732055	CAFE ROCAMORA S.R.L.	RES. 371-AGIP-2010	01/08/2010
1260	1507	9	30501077131	CAFES LA VIRGINIA S.A.	Res. 430-SHYF/01	01/05/2001
1261	1544	0	30523456144	CAFES MUÑO SRL	RES. 371-AGIP-2010	01/08/2010
1262	9479	7	30504312212	CAGNOLI S.A.	RES. 675-AGIP-2009	01/12/2009
1263	14471	0	30708284250	CAIP COMPANIA ARGENTINA DE INVESTIGACIONES FARMACEUTICAS S.A.	RES. 364-AGIP-2016	01/11/2016
1264	15870	0	33500730638	CAIMARI	RES. 364-AGIP-2016	01/11/2016
1265	3749	6	306633205621	CAJA DE SEGUROS SA	Res. 430-SHYF/01	01/05/2001
1266	4749	2	30554475910	CAJA DE VALORES S.A.	Res. 1465-SHYF/01	01/09/2001
1267	3563	6	30701951251	CAJA NOTARIAL COMPLEMENTARIA DE SEGURIDAD SOCIAL	Res. 1465-SHYF/01	01/09/2001
1268	12739	0	30644296875	CALENES SA	RES. 364-AGIP-2016	01/11/2016
1269	11340	0	30500039678	CALEDONIA ARGENTINA COMPAÑIA DE SEGUROS S.A.	RES. 364-AGIP-2016	01/11/2016
1270	13303	0	30683338121	CALENDER SA	RES. 364-AGIP-2016	01/11/2016
1271	9271	3	30555460528	CALERA BUENOS AIRES S.A.	RES. 568-AGIP-2008	01/02/2009
1272	9280	8	30709076902	CALERA CONGRESO SH D E JUAN M SELVA MARIANO F SELVA Y JUAN I SELVA	RES. 568-AGIP-2008	01/02/2009
1273	14607	0	30708555380	CALERA DE SAENZ P.E. S.A.	RES. 364-AGIP-2016	01/11/2016
1274	14152	0	30707587845	CALERA EL OMBU S.A.	RES. 364-AGIP-2016	01/11/2016
1275	16050	0	33683663269	CALERIA SAN JUSTO SA	RES. 364-AGIP-2016	01/11/2016
1276	12253	0	30596774640	CALFOR S.R.L.	RES. 364-AGIP-2016	01/11/2016
1277	10988	4	30710371527	CALFRAC WELL SERVICES (ARGENTINA) SA	RES. 353-AGIP-2014	01/07/2014
1278	15949	0	33596683679	CALICO S.A.	RES. 364-AGIP-2016	01/11/2016
1279	14658	0	30708687568	CALMED SA	RES. 364-AGIP-2016	01/11/2016
1280	16136	0	33707528449	CALOREX S.A.	RES. 364-AGIP-2016	01/11/2016
1281	9089	6	30708046093	CALZADO LOPEZ TABO SRL	RES. 313-AGIP-2008	01/10/2008
1282	11908	0	30551587300	CALZADOS FERLI S.A.	RES. 364-AGIP-2016	01/11/2016
1283	9091	5	30690631721	CALZARTE SRL	RES. 313-AGIP-2008	01/10/2008
1284	12150	0	30585357657	CALZETTA NEUMATICOS SOCIEDAD ANONIMA	RES. 364-AGIP-2016	01/11/2016
1285	10800	4	33690508919	CALZIFICIO ALDAMA SA	RES. 552-AGIP-2012	01/10/2012
1286	9641	8	30700955008	CAMARA DE EMPRESAS DE CONTROL Y ADM. DE INFRAACCIONES DE TRANS DE LA REP. ARG. ASOCIACION CIVIL	RES. 144-AGIP-2010	01/05/2010
1287	10323	5	30643748823	CAMARAS Y LUCES SA	RES. 259-AGIP-2012	01/06/2012
1288	15308	0	30710776063	CAMARIQAS SA	RES. 364-AGIP-2016	01/11/2016
1289	12621	0	30632829732	CAMASA SA	RES. 364-AGIP-2016	01/11/2016
1290	11392	0	30501949891	CAMBRE INDUSTRIAL, COMERCIAL Y FINANCIERA SA	RES. 364-AGIP-2016	01/11/2016

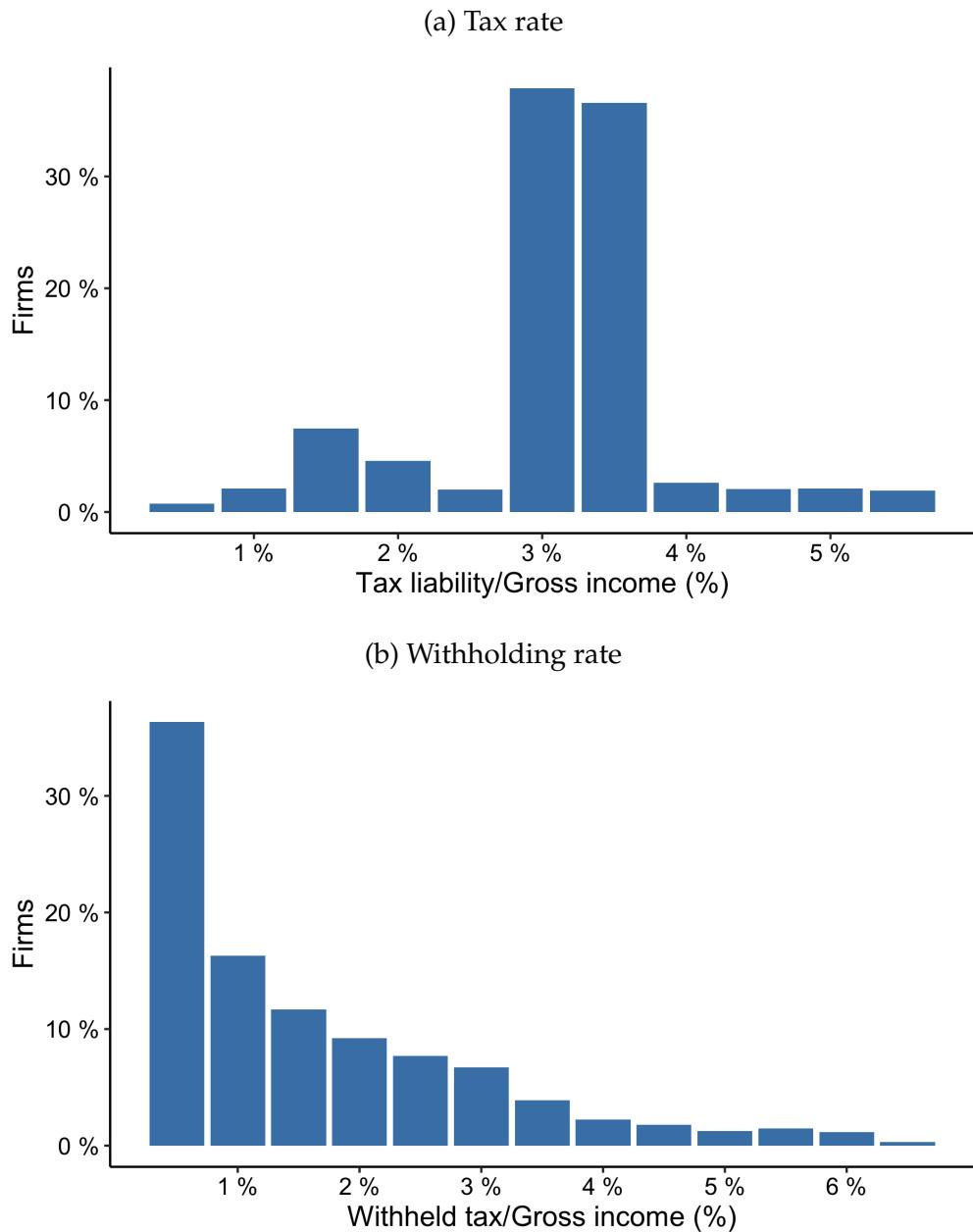
Notes: This figure shows a page from the register of withholding agents published monthly by the tax administration. The full list, in particular, includes all the active withholding agents as of May 2017. This information is publicly available on [AGIP website](#). Each row corresponds to a different firm. The fourth column shows the tax identifier, the fifth column shows the name of the company, the sixth column shows the normative that appoints the firm to act as collection agent, and the last column shows the starting date. Not surprisingly, the majority of the dates correspond to November 1st, 2016, the date in which the major reform we analyze was implemented. We downloaded all these monthly registers, digitized them, and constructed a database that is merged with our main administrative data. We also use this database to construct Figure 3. It is interesting the mix of firms appointed to act as CAs. In this list one can see coffee producers, pharmaceutical firms, insurance companies, footwear manufacturers, construction materials wholesalers, among others.

Figure C.4: Macro evidence that documents the expansion of withholding



Notes: This figure shows macro facts that document the implementation of the reform in the City of Buenos Aires. Panel (a) shows the share of total turnover tax revenue raised via withholding by collection agents. Panel (b) shows the levels of revenue raised by CAs in real terms. Both (a) and (b) display a sharp increase right at the time the net of withholding agents was expanded (denoted with a vertical line). The series keeps increasing during 2017 suggesting some sluggish adjustment of newly appointed firms. Panel (c) shows the percentage difference in tax revenue between the City of Buenos Aires (CABA) and the Province of Buenos Aires (PBA) relative to October 2016, the month before the reform we analyze in the paper. This is analogous to running a difference-in-difference regression between these two jurisdictions and, therefore, captures the aggregate causal effect of appointing firms as tax collectors (i.e., it captures both direct and indirect effects). The graph includes a separate linear fit before and after the reform. It shows that CABA was on a downward tax revenue trend relative to PBA, which is reverted right after CABA expanded their network of collection firms. By 2019, CABA's tax revenue is 20% compared to PBA and relative to October 2016. Source: own elaboration based on aggregate data from 'Dirección General de Estadística y Censos (Ministerio de Hacienda y Finanzas GCBA)'.

Figure C.5: Tax and withholding rates



Notes: The figures show the “effective rates”. Average annual rates by taxpayer calculated at the baseline period, 2016.