

DISCUSSION PAPER SERIES

DP13276

CONSUMERS AS TAX AUDITORS

Joana Naritomi

**DEVELOPMENT ECONOMICS AND
PUBLIC ECONOMICS**



CONSUMERS AS TAX AUDITORS

Joana Naritomi

Discussion Paper DP13276
Published 23 October 2018
Submitted 12 October 2018

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
Tel: +44 (0)20 7183 8801
www.cepr.org

This Discussion Paper is issued under the auspices of the Centre's research programme in **DEVELOPMENT ECONOMICS AND PUBLIC ECONOMICS**. Any opinions expressed here are those of the author(s) and not those of the Centre for Economic Policy Research. Research disseminated by CEPR may include views on policy, but the Centre itself takes no institutional policy positions.

The Centre for Economic Policy Research was established in 1983 as an educational charity, to promote independent analysis and public discussion of open economies and the relations among them. It is pluralist and non-partisan, bringing economic research to bear on the analysis of medium- and long-run policy questions.

These Discussion Papers often represent preliminary or incomplete work, circulated to encourage discussion and comment. Citation and use of such a paper should take account of its provisional character.

Copyright: Joana Naritomi

CONSUMERS AS TAX AUDITORS

Abstract

Access to third-party information trails is widely believed to be critical to the development of modern tax systems, but there is limited direct evidence of the effects of changes in information trails. This paper investigates the enforcement effect of an increased availability of third-party information, and sheds light on how governments can harness this information despite collusion opportunities. I exploit unique administrative data on firms and consumers from an anti-tax evasion program in Sao Paulo, Brazil (Nota Fiscal Paulista) that created monetary rewards for consumers to ensure that firms report final sales transactions, and establishes an online verification system that aids consumers in whistle-blowing firms. Using variation in intensity of exposure to the policy, I estimate that firms' reported revenue increased by at least 21% over four years. Heterogeneous effects across firms shed light on mechanisms: the results are consistent with fixed costs to conceal collusive deals and positive shifts in detection probability from whistle-blower threats. I also investigate the effect of whistle-blowers directly: firms report 7% more receipts and 3% more revenue after receiving the first consumer complaint. To study the role of the value of rewards in improving enforcement, I show evidence consistent with the possibility that lottery incentives amplify consumer responses due to behavioral biases, which would make it more costly for firms to try to match government incentives in a collusive deal. Finally, I find that although firms significantly adjusted reported expenses, there was an increase in tax revenue net of rewards of 9.3%.

JEL Classification: D83, H25, H26, K34, O17

Keywords: N/A

Joana Naritomi - J.Naritomi@lse.ac.uk
LSE

Acknowledgements

I am very grateful to my advisors Alberto Alesina, Raj Chetty, Michael Kremer and Andrei Shleifer. I thank Laura Abramovsky, Tim Besley, Ciro Biderman, Raluca Dragusanu, Itzik Fadlon, Francois Gerard, Nathan Hendren, Nate Hilger, Martin Kanz, Asim Khwaja, Henrik Kleven, Camille Landais, Anders Jensen, Olivia S. Mitchell, Sendhil Mullainathan, Arash Nekoei, Dina Pomeranz, Francisco Queiro, Tristan Reed, Heather Schofield, Edson Severnini, Joel Slemrod, Sandra Sequeira, Ian Tomb, Matthew Weinzierl and Danny Yagan for invaluable comments and advice. I am extremely grateful to Adrea Calabi, Andre Luis Grotti Clemente, Igor Baramboin, and Euripedes De Oliveira for the opportunity to work with the Department of Finance of Sao Paulo (SEFAZ/SP). I am very thankful to Paulo Yamada and to the SEFAZ/SP staff for outstanding collaboration. I thank Sabrina Naritomi and Tiffany Blackman for legal advice, and SOX Consult for assistance with the Brazilian tax law. Research support from the LEAP and Lemann Foundation is gratefully acknowledged. This work uses anonymized data, and does not necessarily reflect the views of SEFAZ/SP.

Consumers as Tax Auditors

Joana Naritomi, London School of Economics¹

October, 2018

Abstract

Access to third-party information trails is widely believed to be critical to the development of modern tax systems, but there is limited direct evidence of the effects of changes in information trails. This paper investigates the enforcement effect of an increased availability of third-party information, and sheds light on how governments can harness this information despite collusion opportunities. I exploit unique administrative data on firms and consumers from an anti-tax evasion program in Sao Paulo, Brazil (*Nota Fiscal Paulista*) that created monetary rewards for consumers to ensure that firms report final sales transactions, and establishes an online verification system that aids consumers in whistle-blowing firms. Using variation in intensity of exposure to the policy, I estimate that firms' reported revenue increased by at least 21% over four years. Heterogeneous effects across firms shed light on mechanisms: the results are consistent with fixed costs to conceal collusive deals and positive shifts in detection probability from whistle-blower threats. I also investigate the effect of whistle-blowers directly: firms report 7% more receipts and 3% more revenue after receiving the first consumer complaint. To study the role of the value of rewards in improving enforcement, I show evidence consistent with the possibility that lottery incentives amplify consumer responses due to behavioral biases, which would make it more costly for firms to try to match government incentives in a collusive deal. Finally, I find that although firms significantly adjusted reported expenses, there was an increase in tax revenue net of rewards of 9.3%.

¹J.Naritomi@lse.ac.uk. I am very grateful to my advisors Alberto Alesina, Raj Chetty, Michael Kremer and Andrei Shleifer. I thank Laura Abramovsky, Tim Besley, Ciro Biderman, Raluca Dragusanu, Itzik Fadlon, François Gerard, Nathan Hendren, Nate Hilger, Martin Kanz, Asim Khwaja, Henrik Kleven, Camille Landais, Anders Jensen, Olivia S. Mitchell, Sendhil Mullainathan, Arash Nekoei, Dina Pomeranz, Francisco Queiro, Tristan Reed, Heather Schofield, Edson Severnini, Joel Slemrod, Sandra Sequeira, Ian Tomb, Matthew Weinzierl and Danny Yagan for invaluable comments and advice. I also thank seminar and conference participants. I am extremely grateful to Adrea Calabi, André Luis Grotti Clemente, Igor Baramboin, and Euripedes De Oliveira for the opportunity to work with the Department of Finance of Sao Paulo (SEFAZ/SP). I am very thankful to Paulo Yamada and to the SEFAZ/SP staff for outstanding collaboration. I thank Sabrina Naritomi and Tiffany Blackman for legal advice, and SOX Consult for assistance with the Brazilian tax law. Research support from the LEAP and Lemann Foundation is gratefully acknowledged. This work uses anonymized data, and does not necessarily reflect the views of SEFAZ/SP.

Tax revenue as a share of GDP is substantially higher in modern advanced economies than in the early 20th century or in present-day developing countries (Besley & Persson 2014). A key source of the variation in tax revenue is the enforcement capacity of governments.¹ In particular, a growing literature emphasizes that information on taxable transactions shared with third-parties can be leveraged by governments to ensure more accurate self-reporting,² and that the increased availability of third-party information trails as countries develop could help explain the dynamics of government revenue among advanced economies during the last century (Gordon & Li 2009; Kleven *et al.* 2016).

Despite the empirical literature on the deterrence effect of third-party reporting, there is little direct evidence on whether changes in availability of information trails can improve compliance and on the mechanisms through which third-party reporting deters evasion, as it hinges on avoiding collusion opportunities among the informed parties.³ This paper exploits quasi-experimental variation and unique administrative data on firms and consumers from an anti-tax evasion program in Sao Paulo, Brazil – *Nota Fiscal Paulista* (NFP) – that created monetary rewards for consumers to ensure that firms report final sales transactions. The program provides tax rebates and monthly lottery prizes for consumers who ask for receipts, and establishes a direct communication channel between the tax authority and consumers through an online account system, where consumers can verify receipts reported by firms and can act as whistle-blowers by filing complaints.

The program was designed to address the ‘last mile’ problem of the self-enforcing mechanism of the Value Added Tax (VAT). Along the supply chain, the tax credit and debit system of the VAT generates third-party reporting in transactions across firms.⁴ At the final consumer stage, however, these self-enforcing incentives break down since consumers typically derive no direct monetary benefit from asking for receipts.⁵ The NFP policy introduced incentives similar to the

¹Musgrave (1969) emphasized the historical relevance of tax administration for tax collection. In the policy debate, tax administration and the enforcement capacity of developing country governments are central issues (Slemrod & Gillitzer 2013; Bird & Gendron 2007; Fund 2011).

²Audit experiments typically detect near zero evasion in income subject to third-party reporting (Kleven *et al.* 2011; Pomeranz 2015). For instance, wage earners would face a much higher risk of audit relative to the self-employed if they under-report income, as firms typically also report wages paid to the government (Slemrod 2007). More generally, information trails shared with third-parties such as employees, suppliers, banks or customers could have a deterrence effect even if they are not systematically reported to the government. Evidence from Denmark and the U.S. suggests that even when income is not subject to systematic third-party reporting, compliance is well below full evasion, which could be explained in part by the existence of derivative information shared with third-parties (Kleven 2014).

³This is a well-known issue in the mechanism design literature (e.g., Tirole 1986): once more than one person is informed about evasion then there are many mechanisms that can be used to elicit that information (Besley & Persson 2013). A key assumption in these cases is that there is no scope for collusion among the informed parties. For instance, Yaniv (1993) argues that employers and employees can find mutually beneficial opportunities to reduce their tax liabilities, which would result in limited enforcement effect on self-reports of individual income subject to cross-reporting by firms.

⁴Most countries in the world adopted the VAT instead of sales tax, perhaps because of its enforcement advantage (Keen & Lockwood 2010). Kopczuk & Slemrod (2006) argue that retail sales tax and the VAT are theoretically equivalent, but the VAT has built-in enforcement incentives along the supply chain. Pomeranz (2015) provides empirical evidence for the self-enforcing properties of the VAT in business-to-business transactions.

⁵Slemrod (2007) refers to the enforcement problem at the final consumer stage as the ‘Achilles heel’ of administering a retail sales tax: if firms collude to underreport transactions, the self-enforcing mechanism can unravel, and may hinder tax collection across the entire chain.

VAT for final sales: it aims to affect both the likelihood that a transaction is reported at all, and the accuracy of reporting, since rewards to consumers are an increasing function of the value of receipts.⁶

I begin by laying out a conceptual framework to discuss how incentives to consumers can affect firm behavior. The NFP policy is effectively increasing the availability of third-party information trails through rewards to consumers, but collusion between consumers and firms could hinder the self-enforcing effect of third-party information. However, in order to collude with consumers and continue evading, firms would need to transfer part of evasion rents to consumers through discounts and incur in a fixed cost to set the collusive deal. Moreover, firms would reveal evasion information to many third parties by conditioning the discount on not accurately reporting the transaction to the government. As in [Kleven *et al.* \(2016\)](#), the difficulty in sustaining collusion with a large number of informed economic agents who can act as whistle-blowers might be important to deter evasion. Therefore, the effect of consumer monitoring should be stronger the higher the threat of whistle-blowing, and the more costly it is for firms to match the value of the rewards offered by the government.

In order to empirically investigate the extent to which rewards to consumers can affect firm compliance, I construct unique administrative data on firm-level monthly tax returns, monthly individual-level data on requested receipts and overall participation in the NFP program, based on administrative records from the tax authority of the state of Sao Paulo.⁷ I divide my analysis into four parts. First, I study the effect of consumer monitoring on firms' compliance by exploiting variation in the intensity of exposure to the policy. I compare reported revenue changes in firms that sell mostly to final consumers (retail) versus firms that sell mostly to other firms (wholesale). I estimate that reported revenue in retail increased on average by 21% over four years as a result of NFP. This estimate is likely to be a lower bound for the effect of the program, given that wholesale firms may also have been affected by the change in consumers' decisions to ask for receipts.⁸

Second, I shed light on mechanisms by examining the implications from the conceptual framework for firms subject to higher whistle-blower threats, and by discussing the role of rewards offered by the government on consumer participation in the enforcement policy. I find evidence consistent with the argument that collusion might be costly and difficult to sustain if consumers can blow the whistle. Firms in sectors that typically have a high volume of transactions and sell small ticket items are effected more, consistent with fixed costs to collude with consumers. Also, firms in sectors that are characterized by a large number of different consumers - that would be more exposed to potential whistle-blowers - are relatively more affected by the consumer rewards program. Furthermore, I link consumer participation to firm compliance by exploiting the timing of consumers' whistle-blowing and find that firms report 7% more receipts and 3% more revenue

⁶Both the tax rebate and the number of lottery tickets with which consumers are rewarded are a function of the total amount they spend in a given month as detailed in Section 2.2.

⁷A number of measures were taken to de-identify the data in order to protect confidential tax records.

⁸Wholesale firms can sell to final consumers directly, in which case the rewards program applies. Additionally, improving compliance among retail firms can affect compliance by wholesalers through the self-enforcing mechanism of the VAT.

after receiving the first complaint.

Third, I turn to the effects of rewards on consumer participation. As suggested by the conceptual framework, the more consumers value the rewards, the more costly it will be for firms to try to match the government's incentives. I exploit variation from lottery prize rewards from NFP to analyze changes in the number of receipts that individuals ask and the total value of receipts. I find that consumers condition their decisions to ask for receipts on past lottery wins. Even when prizes are as small as U.S. \$5, winners ask for receipts more often for at least six months after the lottery result relative to non-winners with the same odds of getting a prize. In addition, I find an increase of U.S. \$16.14 (s.e. 3.08) in the total value of receipts during the 6 months after the lottery, so the effect cannot be attributed to the cash prize alone. The results are consistent with the possibility that lotteries amplify consumer engagement due to behavioral biases.

In the final part of the paper, I discuss implications for tax policy. First, I analyze how tax liabilities were affected by the policy. An increase in reported revenue could, in principle, be circumvented by adjustments in reported expenses. I find that firms significantly increase reported inputs, but not enough to offset the effect of reported revenue on value added. I estimate that the NFP program increased reported tax liabilities by 25.9%. However, through the incentives to consumers, governments are foregoing a fraction of both marginal and infra-marginal revenue, so it is not obvious ex-ante that a positive effect on tax revenue would be sufficient to cover the costs of the program. I calculate that tax revenue increased by 9.3% net of rewards. Finally, I discuss welfare implications of consumer reward policies by considering social costs and benefits, and the impacts on firms, consumers and the government.

This paper contributes to the vast literature on tax enforcement (e.g., [Andreoni et al. 1998](#); [Slemrod & Yitzhaki 2002](#)) by providing evidence on how a policy can tap into local information on tax evasion, and leverage frictions to collusion in order to increase compliance in a hard-to-tax sector. In particular, this paper contributes to a growing literature that argues that third-party information is key for compliance ([Kleven et al. 2011](#); [Pomeranz 2015](#); [Kumler et al. 2012](#)). Additionally, the paper contributes to the literature on the challenges of tax enforcement in developing countries, which is believed to be a key determinant of countries' choices of tax instruments (e.g., [Gordon & Li 2009](#); [Best et al. 2015](#); [Jensen 2016](#)). In particular, a growing strand of the literature sets aside non-compliance due to firm non registration at the tax authority – the formal-informal margin – and instead examines non-compliance among formal firms.⁹

The paper also contributes to the policy debate on sales tax enforcement. Many countries adopted policies to reward consumers to address the last mile problem of the VAT.¹⁰ This paper

⁹[De Paula & Scheinkman \(2010\)](#) argue that better enforcement in the intensive margin of VAT systems can endogenously generate incentives to formalize by creating supply chains of formal firms. See [Bruhn & McKenzie \(2014\)](#) for a review of the literature on the formalization of firms.

¹⁰For instance, Argentina, Bolivia, Brazil, China, Chile, Colombia, Indonesia, Italy, Portugal, Puerto Rico, South Korea and Slovakia, among other countries, have introduced policies to address the enforcement problem downstream through monetary incentives – through tax refunds, lotteries, or fines – for consumers to request receipts ([Fookien et al. 2015](#); [Bird 1992](#); [Cowell 2004](#); [Fabbri 2015](#); [Marchese 2009](#)). [Wan \(2010\)](#) argues that a program that turns receipts into lottery tickets in China was effective in raising tax revenue, but the evidence for such policies is mixed ([Barroso & Cortez 2007](#); [Mattos et al. 2013](#)).

provides, to my knowledge, the first direct evidence of consumer behavioral responses to rewards from asking for receipts. The results also reinforce existing findings on individual responses to lotteries that are used as levers in other contexts, such as lottery-linked savings (Tufano 2008; Kearney *et al.* 2010). Moreover, the evidence from the NFP lotteries adds to the literature on the behavioral effects of lottery wins such as the lucky store effect (Guryan & Kearney 2008). More generally, the paper sheds light on the effects of participatory policies used as a monitoring tool.

Finally, the paper contributes to the debate on how other margins of adjustment (e.g, reported expenses) may compensate for an increase in enforcement of revenue reporting. In the context of corporate income tax (CIT), Carrillo *et al.* (2017) and Slemrod *et al.* (2017) find that reported costs substantially increase, offsetting to a large extent the profit change from more accurate revenue reporting. In both cases, the cost increase occurred primarily in difficult to verify margins such as “Other expenses”. In the case of the VAT, the ability to adjust inputs is arguably relatively more limited as a tax credit must be another firms’ tax debit. The findings of this paper indicate that reported expenses can be adjusted in a VAT context, but perhaps to a lesser extent than in CIT due to the self-enforcing feature of the VAT.

The remainder of the paper is organized as follows. Section 1 outlines a simple conceptual framework to guide the empirical analysis. Section 2 describes the institutional background of the *Nota Fiscal Paulista* program, the relevant datasets, sample definitions and summary statistics. Section 3 investigates the enforcement effect of the introduction of third-party information through consumer rewards on firms’s reported revenue, and Section 4 sheds light on mechanisms suggested by the conceptual framework regarding whistle-blower threats, collusion costs and the value of monetary rewards. Section 5 examines the impact on expenses and tax revenue, and discusses implications for tax policy. Section 6 concludes.

1 Conceptual framework

I begin the analysis by describing a simple conceptual framework that examines the degree to which consumer monitoring can affect the evasion decision by firms. I use a variant of the Allingham & Sandmo (1972) framework discussed by Kleven *et al.* (2011), in which the probability that a taxpayer is caught evading depends on the audit rate and the probability of detection conditional on an audit. First, I present a baseline case with government monitoring only. Then, I introduce consumer monitoring as an additional enforcement tool that gives rewards for consumers to ensure firms report final sales transactions, and allows consumers to act as whistle-blowers. In this case, firms may continue evading by colluding with consumers. In doing so, however, firms get a lower benefit from evasion and reveal to a number of third-parties evasion information that the government may access through whistle-blowers. As a result, firms will increase compliance.¹¹

¹¹I take a positive approach to understand the effects of different monitoring tools on firms evasion decision. For a normative approach, see Arbex & Mattos (2014) that investigate how the Ramsey equation is modified once consumers are rewarded to ask for receipts.

1.1 A tax evasion model

Consider a risk-neutral firm that pays a tax $\tau \in [0, 1]$ proportional to their reported revenue $Y \geq 0$. Suppose that firms sell a single product, and that each firm has N consumers who each make one purchase which generates revenue $\bar{y} \geq 0$. Firms have a true pre-tax revenue $\bar{Y} = N\bar{y}$, and choose to report revenue Y to maximize profits π , which depends on enforcement policies.

1.1.1 Government monitoring only

Let $p \in [0, 1]$ be the probability of detection faced by the firm. Similarly to [Kleven *et al.* \(2011\)](#), I assume that the audit probability is increasing in the amount evaded, and that the probability governments detect evasion is a product of government audits and the likelihood that the government will uncover evasion by taxpayers during an audit. [Kleven \(2014\)](#) argues that the more derivative information from various third party sources is available to tax enforcement, the more compliance we observe despite low audit rates. The intuition is that the more information the government has about the firm, the easier it is to detect evasion conditional on an audit ([Slemrod 2007](#)).

Let $a(E) \in [0, 1]$ be the audit probability, $E = \bar{Y} - Y$ be the total evasion by firm, and $d \in [0, 1]$ be the ability of the government to detect evasion in an audit. The probability of detection faced by the firm can be written as $p \equiv a(E)d$, $p'(E) = a'(E)d > 0$. If the firm is caught evading, the government applies a fine $\theta \geq 0$ in proportion to the evaded tax $\tau(\bar{Y} - Y)$. For simplicity, assume that in the absence of monetary incentives, consumers do not ask for receipts and have no impact on the evasion decision of firms. Thus, firms report revenue Y to maximize:

$$\pi = (\bar{Y} - \tau Y)(1 - p) + [(1 - \tau)\bar{Y} - \theta\tau(\bar{Y} - Y)]p \quad (1)$$

An interior optimal solution Y^* satisfies the first order condition $d\pi/dY = 0$:¹²

$$[a(E) + a'(E).E]d(1 + \theta) = 1 \quad (2)$$

The right hand side of equation (2) is the marginal benefit of evading an extra dollar, and the left-hand side is the marginal cost of evading that extra dollar. As discussed in [Kleven *et al.* \(2011\)](#), the firm that evades an extra dollar incurs in a higher probability of audit of all infra-marginal dollars evaded. Firms choose the optimal Y^* that satisfies equation (2), and Y^* will be increasing in the detection probability d .¹³

¹²The second-order condition is $2a'(E) + a''(E)E > 0$. It is necessary and sufficient that $Ea(E)$ is convex, which is essentially the same condition as in [Kleven *et al.* \(2011\)](#).

¹³Given the convexity of $Ea(E)$, $\frac{\partial Y^*}{\partial d} = \frac{[a(E) + a'(E).E](1 + \theta)}{2a'(E) + a''(E)E} > 0$.

1.1.2 Adding consumer monitoring

For simplicity, I assume that if a firm issues a receipt, that transaction will be reported correctly.¹⁴ Now, suppose the government creates targeted incentives for consumers to ask for receipts. Consider the case where consumers are rewarded with $\alpha \in [0, 1]$ of the tax τ firms pay on the transaction reported to the government. Consumers can ensure they receive this reward by requesting a receipt, and they can act as whistle-blowers by informing the government about firms' non-compliance.

Let $\kappa(\alpha)$ be how much consumers value the rewards α from the program, where $\kappa(\alpha) \geq 0$ and $\kappa'(\alpha) \geq 0$. Let $\kappa(0) = 0$, in which case we are back to government monitoring only.¹⁵ For instance, if consumers are unaware of the program, $\kappa(\alpha)$ can be close to zero even if $\alpha > 0$. If consumers enjoy participating in the program above and beyond the monetary value attached to it, $\kappa(\alpha) > \alpha$. This could occur, for instance, if consumers enjoy playing the lottery or value engaging in a tax compliance program. There could also be a framing effect from rewards: a tax rebate, for instance, may help segregate small gains, making it more valuable than a discount of the same amount (Thaler 1999).¹⁶

As the government is rewarding consumers with a fraction of what firms pay in taxes, firms and consumers could potentially agree to a mutually-beneficial deal and not issue receipts.¹⁷ For simplicity, assume that firms make a take it or leave it discount offer to consumers to continue reporting y instead of the true amount \bar{y} , and that consumers accept a discount deal that matches the difference with respect to the government's reward $\kappa(\alpha)\tau(\bar{y} - y)$.

It is important to note that not only must the firm share part of their evasion rents with the consumer, the firm reveals to a third party that it evades taxes by conditioning the discount on not reporting the true amount of the transaction. Consumers, therefore, become informed third-parties. As consumers can act as whistle-blowers, governments may gain access to relevant information about firms' evasion. Thus, the firm might face an increased detection probability if consumers

¹⁴Receipts are an important tool for enforcement, which is the main rationale for consumer rewards programs in the first place. Therefore, I will refer to firms' decision to issue a receipt as being equivalent to a decision to report a transaction. It is possible that even after issuing a receipt, firms may try to erase those records (using zappers or phantom ware, for instance), but it is costly. In the Sao Paulo context, billing machines that issue receipts with time stamps and serial codes were already widespread since early 2000's. Across the world, the adoption of Electronic Billing Machines that send data in real time to governments are making such expost changes quite difficult.

¹⁵If $\kappa(0) > 0$, i.e., consumers value compliance per se, the policy could have an effect even if there are no monetary rewards, just a whistle-blower channel. In this case, the problem of the firm would be similar as it would try to match this taste for compliance with a discount, and providing a tool to whistle-blow firms could be valuable to create credible threats of $\varepsilon > 0$.

¹⁶Based on prospect theory, framing outcomes separately could yield larger gains as the gain function is steepest at the origin. The utility of a small gain, therefore, could exceed the utility of slightly reducing a large loss.

¹⁷Tirole (1986) discusses the collusion problem in auditing contracts in which a group of informed parties – the auditor and the agent – can manipulate the information reported to the principal. This context is also similar to the case of corruption with theft in Shleifer & Vishny (1993).

cannot commit not to whistle-blow.¹⁸

Kleven (2014) argue that a key deterrent of collusion is the sheer number of internal or external parties to which a firm that evades taxes exposes itself. I consider the case where there is a positive probability of a random shock between the parties that can trigger a consumer to blow the whistle. A random shock could be generated by some conflict between the consumer and the shopkeeper, or a moral concern of the consumer. Therefore, the larger the number of consumers N , the higher the additional risk of detection introduced by consumers that may act as whistle-blowers.

Assume $\varepsilon > 0$ is the probability that such a random shock occurs; let ε be i.i.d. across consumers. Assume that if one consumer blows the whistle on the firm the information she provides allows the government to detect evasion with certainty in an audit, and that all the N consumers may blow the whistle. The intuition is that consumers are gathering relevant information about evasion conducted by firms that can improve the enforcement capacity of the government for a given audit rate. So the ability of the government to detect evasion under consumer monitoring will be a function of N and can be written as $d_c(N) = 1 - (1 - d)(1 - \varepsilon)^N \geq d$, $d'_c(N) > 0$. Therefore, firms face an increased probability of getting caught p_c given by $p_c = a(E)[1 - (1 - d)(1 - \varepsilon)^N]$.¹⁹

Finally, suppose that when firms make an offer to a consumer, they pay a fixed cost $\rho > 0$. The fixed cost can be thought of as a concealment cost paid to reach a collusive deal in each transaction. I assume that, if firms evade taxes, they adopt a collusion policy that applies to all transactions and must pay $\rho > 0$ as concealment cost. In Online Appendix C I discuss an alternative case in which firms only collude with a fraction of the transactions. The main point of collusion costs is to emphasize how frictions from concealing a collusive deal with a large number of parties can contribute to increase compliance.

Now firms choose Y to maximize:²⁰

$$\pi = (\bar{Y} - \tau Y)(1 - p_c) + [(1 - \tau)\bar{Y} - \theta\tau(\bar{Y} - Y)]p_c - \kappa(\alpha)\tau(\bar{Y} - Y) - \rho N \quad (3)$$

As mentioned above, under the new policy, firms have to transfer part of the evasion rents to consumers through discounts. An interior optimal solution Y^{**} satisfies the first order condition $d\pi/dY = 0$:

¹⁸In the empirical context, it is particularly salient that a collusive deal will allow the firm to evade taxes since the government is giving a reward for consumers to ask for receipts in a campaign against tax evasion. Also, as will be described in detail in Section 2, consumers can be whistle blowers by filing complaints about specific firms to the government through a website.

¹⁹It is possible that whistle-blowers affect the audit probability as well, and an alternative model could be written in line with (Kleven *et al.*, 2016) that assumes that one whistle-blower triggers a full audit. The empirical implications in the next sections would be similar. Because information about audit rates or audit strategies are strictly confidential, it is not possible to distinguish in the data changes in audit rate from changes in detection ability conditional on audit. The conceptual distinction is useful, nonetheless, to illustrate how information from consumer monitoring can augment the effectiveness of government audits even if there are no changes in audit rates, which implies a higher risk of evasion faced by taxpayers.

²⁰I assume that if the firm is audited the government will consider as tax evasion the amount not reported based on the posted price \bar{y} , not the discounted price. Therefore, \bar{Y} will be the true revenue of the firm, instead of the revenue net of transfers to consumers.

$$[a + a'(E).E]d_c(1 + \theta) = 1 - \kappa(\alpha) \quad (4)$$

The equation highlights two ways in which the firm's evasion decision is affected by rewards to consumers. First, the marginal benefit of evading an extra dollar is reduced by $\kappa(\alpha)$. Therefore, the degree to which consumers value the program enter as an extra penalty for each dollar evaded. In this case, the more consumers value the rewards α , the higher this extra-penalty will be. Second, if consumers cannot commit not to whistle-blow, the new detection probability will be increasing in the number of consumers N as it increases the chances that consumers will inform the government about the evasion activity of firms.

The cost of collusion ρN would affect the extensive margin decision between evasion and full compliance, but not the intensive margin of compliance. In particular, if the payoff of full compliance and the collusion cost $(1 - \tau)\bar{Y} + \rho N$ is larger than the expected value of taking the evasion gamble, the firm would shift to full compliance.²¹ Note that the value of the transaction \bar{y} will matter for this policy because firms with the same true revenue \bar{Y} may be affected differently: firms that sell small ticket items (low \bar{y}) would have to collude in a larger number of transactions for a given total revenue \bar{Y} . This effect should be empirically similar to the increased probability of getting caught evading through whistle-blowers, although the whistle-blower threat depends on the total number of different consumers, whereas the collusion cost effect depends on the sheer number of transactions as I discuss below.

1.1.3 Comparative statics and Discussion

The first order conditions in both (2) and (4) can be expressed as $a + a'(E).E = c$, where c is a function of parameters that are changing with the policy. With government monitoring only, $c = \frac{1}{d(1+\theta)}$. With consumer monitoring, $c = \frac{1-\kappa(\alpha)}{d_c(1+\theta)}$. Changes in c translate into comparative statics of E^* : if c increases, the optimal evasion will also increase.²² In this subsection, I discuss how each component of the policy affects the evasion decision of firms and its implications.

Value of Rewards In a collusive deal, firms try to match the value of rewards provided by the government through a discount. The reward to consumers reduces c and, therefore, decreases evasion. In particular, the higher the reward α and the more consumers value the rewards $\kappa(\alpha)$, the higher the extra-penalty per dollar evaded will be. In the empirical setting, the reward has a lottery component so it is possible that its value is actually higher than the monetary expected value of the program's reward ($\kappa(\alpha) > \alpha$). As I discuss in Section 4.2.2, a taste for gambling or behavioral biases in assessing the odds of winning prizes could inflate the perceived value of lottery rewards, making it particularly costly for a firm to replicate the government's reward

²¹In Online Appendix C, I consider an alternative model in which firms can selectively collude with some consumer but not others to minimize such fixed costs. In this case, the fixed cost will affect the intensive margin decision of firms to evade taxes. The key insights, however, can be achieved with this simpler version where firms have a collusion policy that applies to all transactions.

²²The necessary assumption for monotonicity $\frac{\partial E^*}{\partial c} = \frac{1}{2a'(E)+a''(E)E} > 0$ is that $Ea(E)$ is convex, which is the same convexity assumption discussed above for the second order condition.

through a discount. The rewards, therefore, could have an effect on compliance even without a whistle-blower channel (i.e., unchanged detection ability of the government).

Further, the fact that the valuation of the benefit may differ from the monetary rewards could introduce a higher fixed cost in reaching a collusive deal. In the simple framework considered above, a larger ρ could push firms to switch to compliance. Considering a simple extension where the fixed cost ρ for colluding with consumers increases with the gap between the perceived size of the reward and its monetary value, i.e, if $\rho(\kappa(\alpha) - \alpha)$ and $\rho'(\kappa(\alpha) - \alpha) > 0$, it could be the case that a reward scheme that is not straightforward for a firm to mimic (e.g., lottery) is relatively more cost-effective for the government. The impact on consumers' welfare, however, depends on why there is such bias in the perceived value of rewards, and whether it affects 'experienced' utility or only 'decision' utility. I revisit this discussion in Section 5.

Volume of consumers. The enforcement change introduced by consumer monitoring is stronger the larger the increase in the detection probability. This comparative statics follows from a drop in c as $d_c > d$. The increase in compliance should increase with the volume of consumers for a given firm size or true revenue \bar{Y} . This distinction between firm size and volume of consumers is relevant to shed light on a mechanism through which third-party information affect compliance: exposure to whistle-blower threats can decrease evasion.

Volume of transactions and size of transactions In the simple framework proposed above, there is one transaction for each consumer. Considering the case of multiple transactions per consumer, the number of transactions could matter through a different mechanism than the number of consumers: the larger the number of transactions a firm has, the more this policy may increase compliance as the fixed costs dis-proportionally affect firms that need to collude multiple times for a given firm size (ρN). Similarly, firms that sell small ticket items should be affected relatively more for a give firm size.

Whistle-blowers. If firms engage in a collusive deal but believe that consumers will likely never whistle-blow - perhaps because they have never seen a consumer blow the whistle - they may perceive ε as being lower than it actually is. If this is the case, once firms observe consumers blowing the whistle, they may update upwards their beliefs about ε , which reduces c and increase compliance. The data allows me to directly observe how firms react once they learn a consumer blew the whistle.

Firm size. Even absent of consumer monitoring, larger firms arguably would be more at risk of getting caught for a given evasion level due to the number of internal or external informed parties (e.g., employees, buyers, suppliers) that may have information on tax evasion. In the model above, it would be as if d was higher for larger firms to begin with and, therefore, their baseline level of compliance should be higher (higher d lowers c). In other words, firm size alone could already generate variation in baseline exposure to whistle-blower threats (e.g., from the total number of informed third parties that firms interact with). The consumer rewards policy introduces a new set of informed third-parties: consumers that will ask for receipts and potentially learn about evasion (e.g., if they observe collusion through conditional discounts offers). In this case, the effect of the

policy would be relatively larger for firms with a lower baseline d , which can be tested in the data.

Government Revenue. The government transfers to consumers $\alpha\tau Y^{**}$. It is important to notice that because the government cannot distinguish between marginal and infra-marginal sales, it rewards infra-marginal sales as well. If we restrict attention to revenue from taxes only (without fines), the program should increase tax revenue if $Y^{**} - Y^* > \alpha Y^{**}$. In other words, government revenue will increase if the tax base increase is larger than the tax base the government is forgoing. This implies that the percentage change in tax revenue will have to be at least $\frac{Y^{**}-Y^*}{Y^*} > \frac{\alpha}{(1-\alpha)}$ to generate an increase in government revenue. If the baseline level of compliance is very low, the program would be particularly attractive. Also, if higher compliance can be achieved without rewarding every single transaction (e.g., imperfect take up by consumers), the program will be more cost-effective. I revisit this discussion in Section 5.²³

2 Institutional Background and Data

This Section provides institutional background on the *Nota Fiscal Paulista* (NFP) policy, and the details of the program that are important for the empirical analysis. First, I briefly introduce the relevant features of the Brazilian tax system and the NFP policy. Then, I describe the datasets I use and sample definitions.²⁴

2.1 Institutional Background

The State of Sao Paulo is the largest state in Brazil: it accounts for 34% of the country's GDP, and has a population of 42 million people. The metropolitan area of Sao Paulo is the second most populous in the Americas. The state of Sao Paulo depends mostly on its own tax revenue, as opposed to federal transfers.²⁵ States in Brazil have two main tax instruments: a tax on goods and certain services (ICMS) and a property tax on motor vehicles (IPVA).²⁶ The ICMS is a value added tax (VAT), and it is the most important source of revenue in Sao Paulo. Because the ICMS is a state-level tax in Brazil, its legislation and enforcement policies are determined by the states. The tax base includes goods and some services, the most common ICMS rate is 18% over the valued

²³Note that, for simplicity, I am assuming here that the increase in tax collection is proportional to the increase in reported revenue as it is modelled as a sales tax. The goal of this simplification is to flesh out how sales reporting - the margin directly affected by the policy - can change. The same logic would carry over to the VAT if the tax base Y is the reported value added, and changes in reported expenses cannot fully offset changes in reported revenue. This is true in the data as I analyze in Section 5.

²⁴Throughout the paper I will convert Brazilian Reais to dollars using U.S.\$1=R\$2 exchange rate, which is the average exchange rate during the period of analysis (2004 – 2011).

²⁵When the NFP policy was implemented in 2007, Sao Paulo's own tax revenue was 75% of its total revenue according to the balance sheets of the Brazilian Treasury Department. Moreover, Haddad *et al.* (2011) argue that Sao Paulo state generated more than 40% of the Federal tax revenue, while receiving less than 35% of Federal transfers in 2005. Federal taxes include, for instance, individual and corporate income taxes, payroll taxes and taxes on manufactured products.

²⁶The IPVA ("Imposto sobre Propriedade de Veículos Automotores") and ICMS ("Imposto sobre Circulação de Mercadorias e Serviços") typically account for 95% of the total tax collected by states. The other two sources of tax revenue are a tax on bequests and donations called ITCMD ("Imposto sobre Transmissão Causa Mortis e Doações") and fees for public services.

added, which is computed through a credit-invoice method.²⁷ As is common in VAT across the world, there is a threshold below which firms pay taxes on gross revenue instead of the value added (Keen & Mintz 2004). Firms that have yearly gross revenue of less than U.S. \$1.2 million can choose to be in a simplified tax regime called SIMPLES in which firms pay taxes based on gross revenue. The ICMS average rate in the SIMPLES is 3.5% of gross revenue.²⁸ The majority of the tax collected in retail comes from VAT firms (over 85%).

In 2007 the state of Sao Paulo collected U.S. \$27.2 billion with the ICMS, equivalent to 7.6% of the state's GDP. Overall, tax revenue in Brazil is high for developing country standards. Considering all taxes, tax revenue amounts to 34% of the country's GDP (Fund 2011). Nonetheless, there are many reasons to believe that tax compliance is not perfect in Brazil. According to La Porta & Shleifer (2014), estimates of size of the country's informal economy range from 19% to 34% of GDP. Unregistered firms are invisible to the tax authority, and no taxes are levied directly on them. Formal firms have to report their activity to the tax authority on a monthly basis, and pay the ICMS in relation to their reported activity. Despite the tax authority's monitoring, compliance by formal firms is also limited. In the World Business Environment Survey 2003, on average Brazilian formal firms claim that 20-30% of sales are not reported to the tax authority by a typical firm in their area of activity.²⁹ When the NFP program was implemented, the Secretary of Finance of Sao Paulo at the time argued that the retail sector in the state evaded taxes on approximately 60% of its sales (Jornal Estado Sep.2007).

2.2 The Nota Fiscal Paulista program

The *Nota Fiscal Paulista* (NFP) program was created by the government of the state of Sao Paulo in October 2007 in order to reduce tax evasion of the state's VAT, and to foster a culture of tax compliance.³⁰ The idea behind the NFP program is to use consumers as "tax auditors" by introducing targeted incentives for consumers to ensure that firms report final sales. The incentives provided by the program replicate the VAT self-enforcement already in place for business to business transactions; rewards are increasing in the value of the purchase such that buyers have incentives to ask for receipts, and to make sure that the value of the purchase is reported correctly by the supplier. Therefore, the NFP program directly affects two forms of under-reporting: (i) firms may not report

²⁷On a monthly basis, firms declare how much taxes they owe based on their taxable sales (debits) and how much credit they have from purchases of goods and services taxed by ICMS (credits). The difference between tax debits and credits determine a firm's tax liability. For the majority of goods, the ICMS rate is 18%. In some cases, a reduced rate of 7% or a higher rate of 25% is applied. The tax base covers goods and a few services, such as restaurants and electricity provision. Most services are part of the tax base of a municipal sales tax.

²⁸For more details about SIMPLES see De Paula & Scheinkman (2010) or Monteiro & Assunção (2012).

²⁹The question in Batra *et al.* (2003) is: "Recognizing the difficulties many enterprises face in fully complying with taxes and regulations, what percentage of the total sales would you estimate a typical firm in your area of activity keeps off the books: 1 (none); 2(1-10%); 3 (11-20%); 4 (21-30%); 5 (31 - 40%); 6 (41 -50%); 7 (over 50%)." In the case of firms that sell to final consumers, the tax evasion problem is likely to be more severe since firms are smaller than in upstream sectors. In the same survey, the percentage of sales that are underreported or not reported at all reaches 30-40% among smaller firms in Brazil.

³⁰The NFP policy was framed as an incentive to improve tax morale. The official slogan of the policy was "Incentive Program for Fiscal Citizenship" ("Programa de Incentivo à Cidadania Fiscal").

a transaction at all, or (ii) firms may falsely claim a lower transaction value.³¹

In a nutshell, the program introduced the possibility of identifying an individual taxpayer number – hereinafter referred to as Social Security Number (SSN) equivalents – on each receipt, and created a system of tax rebates and monthly lotteries so that final consumers have incentives to request receipts with their SSN. Since the process of reporting receipts to the tax authority is done by firms, and the consumer’s SSN is attached to it, consumers do not need to send their receipts to the tax authority to get the rewards, which markedly reduces consumer participation costs. Consumers have to create an online account at the tax authority’s website, which allows them to collect rewards and cross-check the receipts issued with their SSNs. The online system also allows consumers to file complaints about specific firms, which introduces a threat that consumers may act as whistle-blowers (see Figure A1 in the Online Appendix).

Implementation. The reward system was introduced along with a system of transaction reporting through which firms were required to send electronically to the tax authority all receipts they issue - with or without a SSN. Previously, firms only reported monthly aggregated information and were required to keep all the supporting documents and receipts in their books. With the new system, firms were also required to send the government individual sales information on a monthly basis.³² Importantly, this new system did not change the technology of receipts issued by retail firms during the period of analysis (2004 -2011). Billing machines that issue receipts with time stamps and serial codes (called *Cupom Fiscal* in Brazil) were already widespread in Sao Paulo in the early 2000’s, and billing machines with real time electronic transmission of receipts directly to the tax authority (called *Nota Fiscal de Consumidor Eletrônica* in Brazil) – that can be an important step to reduce tax evasion (see [Eissa et al. \(2014\)](#)) – only started being introduced in Sao Paulo in 2015. Therefore, the requirement to send disaggregated sales information alone – without incentives to consumers – should not change firm compliance behavior: firms could simply send to the tax authority information on individual transactions they were already reporting in their official books in the end of each month.³³

Eligibility. The government leveraged the fact that SSNs are not considered sensitive information in Brazil.³⁴ Any person that holds a Brazilian SSN equivalent is eligible to participate in the program.³⁵ No pre-registration is needed for consumers to be eligible for tax rebates, but con-

³¹A common way to evade taxes in Brazil is to underreport the value of a sale. This type of evasion is informally known as “meia-nota” or “half-receipt” ([do Amaral et al. 2010](#)).

³²The system is called TD-REDF (“Transmissor de Dados para o Registro Eletrônico de Documento Fiscal” or “Data Transmitter for Electronic Registration of Fiscal Document”).

³³In addition, during the period of analysis product codes and bar codes were not standardized so the itemized information was not used beyond the total value reported in each receipt that was used to calculate the rewards.

³⁴For instance, the Brazilian SSN equivalent (CPF) is written on checks under the signature line, and consumers are frequently asked for their CPF in business transactions. Also, Brazilians have multiple identification numbers that make identity theft more costly: individual identification, taxpayer number, a voter identification, a social security identification, among others.

³⁵Throughout the paper I will refer to the CPF (“Cadastro de Pessoa Física”) as SSN. I will focus on CPF holders only. They are the overwhelming majority of participants in the program. Some NFP participants have a CNPJ (“Cadastro Nacional de Pessoa Jurídica”), which is a SSN for firms. Charitable institutions and condominiums, for instance, also have CNPJ and receive the exact same benefits as final consumers.

sumers must create an online account at the tax authority’s website to be rewarded with lottery tickets for monthly cash prizes.

The reward system. At the moment of purchase, the consumer may ask for the receipt, and give the cashier her SSN. Firms must send all receipts – with or without SSNs – to the tax authority on a monthly basis. As the tax authority receives the receipts, it creates an account for each SSN where it stores all receipt information and the tax rebates due from each receipt. If the consumer has an online account and has opted in for lotteries, the system also automatically generates a lottery ticket for every total of U.S. \$50 spent. Therefore, it does not matter if the U.S. \$50 come from one receipt or 50 receipts of \$1, which also means that the lottery tickets are not attached to specific purchases or shops. During the registration, a consumer may also opt to receive an email every time a receipt is issued with her SSN. The online account displays how much consumers are rewarded for each transaction, and has tabs where a consumer can click to manage rewards and file complaints. In the Online Appendix A, Figure A1a shows an online account example, and Figure A1b displays a receipt with a consumer’s SSN.

Tax rebates. For a given receipt, consumers receive a tax rebate of 30% of the VAT paid by the final sale establishment in a month – i.e., it is only a share of the taxes paid by the retailer, not the total tax collected along the supply chain –, shared among all consumers of that establishment who provided their SSN that month in proportion to their expenditure in that establishment and month. The calculation of the benefit, thus, is a function of an entire month’s worth of SSN receipts and resultant tax paid by the final sales establishment.³⁶ On average, the tax rebate is around 1% of the total value of the purchase.³⁷

Lotteries. NFP has held monthly lotteries since December 2008. For every U.S. \$50 a consumer spends in NFP receipts per month, she receives one lottery ticket.³⁸ If the consumer opts in for these lotteries while enrolling online, lottery tickets are automatically generated based on the consumer’s total expenditures in NFP receipts as described above. Lottery draws are held around the 15th of each month, and each month 1.5 million prizes are distributed on average. Most prizes range from U.S. \$5 to U.S. \$25, and there are usually three large prizes from U.S. \$15,000 to U.S. \$500,000. On average, the expected value of a lottery ticket is 0.2% of the total purchase.

Collecting rewards. Rewards can be: (i) directly deposited into the consumer’s bank account,

³⁶If two consumers buy the same total value in the same shop and month they will receive the same tax rebate even if they bought different goods that may be taxed differently. This is partially due to the fact that the tax authority can only use the total value of the receipt to calculate the rebate shares of each individual since the itemized information in the receipt was not standardized in the period of analysis. More precisely, if the firm has N consumers in a month, the benefit consumer i receives from an NFP receipt depends directly on the total ICMS collected from establishment e in month m ($ICMS_{em}^{total}$), the total value of NFP purchases associated with consumer i and establishment e in month m (V_{iem}) and inversely on the total value of NFP purchases in establishment e in month m ($\sum_{j=1}^N V_{jem}$). Also, there is a cap on how much an individual consumer can receive: 7.5% of the total expenditure, which is 30% of the highest VAT rate (of 25%). Thus, $TaxRebate_{ime} = \min\{0.3 \cdot [ICMS_{em}^{total} \times \frac{V_{iem}}{\sum_{j=1}^N V_{jem}}], 0.075 \cdot V_{iem}\}$.

³⁷The rebate value is 30% of the tax collected from the shop that sold the item. This is consistent with the fact that the value of taxes paid is 4% of the revenue on average. Therefore, for every dollar that she spends, she will get around 1% of cash rebate.

³⁸The lottery draw in month m uses lottery tickets generated by expenditures in month $m - 4$. This 4-month gap is necessary in order to make sure that all disputes over missing or incorrect receipts are resolved before the lottery.

(ii) used to pay other state taxes, (iii) transferred to another person with an online account or to a charity. Consumers must have an online account to manage the rewards. Tax rebates are disbursed biannually. In April, tax rebates from July to December of the previous year are made available to consumers; in October the tax authority disburses tax rebates from purchases between January and June of the same year. Lottery prizes can be collected soon after the results are released. Consumers have up to five years to claim the benefits.

Complaints. Consumers may file complaints regarding a purchase made at a specific establishment up to the 15th of the month following the purchase. The consumer must identify the establishment and select a reason for the complaint from a 5-option menu: (i) the establishment did not issue a receipt; (ii) the establishment refused to write the consumer's SSN on the receipt; (iii) the establishment issued the receipt but did not register it electronically; (iv) there is a discrepancy between the information on the receipt issued to the consumer and the receipt registered electronically at the tax authority; and (v) other reasons.³⁹ Consumers receive a part of the fines paid by the firm as rewards instead of the usual monetary reward when they file a complaint that escalates to a fine. I do not observe, however, the consequences of a given complaint. In the empirical analysis, I therefore use all complaints.

Fines. Firms that do not issue the NFP receipt correctly are subject to penalties and potentially more comprehensive audits by the tax authority. Under tax law, firms can pay up to 100% of the evaded tax, and there are additional penalties for misreporting documents and receipts.⁴⁰ If a firm issues a receipt with an individual SSN and misreports the transaction, the process of punishing firms is straightforward if the consumer has a SSN receipt as proof of purchase.⁴¹ In this case, there are fines applied by the consumer's protection bureau PROCON (Fundação de Proteção e Defesa do Consumidor).

Timeline. NFP was implemented in the retail sector between October 2007 and December 2008. The tax rebate system and electronic submission of receipts was phased-in by groups of sectors between October 2007 and May 2008. The online system to file complaints was available starting in October 2008; the first lottery draw was in December 2008. In April 2009, the tax authority disbursed tax rebates for the first time from all purchases since October 2007, and every 6-months thereafter the government disbursed tax rebates according to the schedule described above.

During the period of analysis from October 2007 to December 2011, 13 million people enrolled online at the tax authority's website, which is 40% of the people ages 15 and above in the state. In a given month there are typically 5 million more people asking for SSN receipts than there are online accounts. This gap highlights that the cost to start participating in the program is relatively small:

³⁹At that point the consumer does not need to provide evidence to support her complaint, and she can describe details of her case in a text box. The establishment is notified that a complaint was filed via email or letter, and it has 10 days to respond to the complaint. If the consumer is not satisfied with the response, she can file an official complaint. Before this point, the tax authority is not involved in the case. If the consumer decides to file an official complaint, she has to submit supporting evidence by scanning or taking a picture of the receipt or any other proof of purchase. From that point onward, the tax authority and the Consumer Protection Bureau will review the case and apply fines accordingly.

⁴⁰For the legislation on tax penalties *Part IV "violations concerning fiscal documents and tax forms"* of Decree 45490/00.

⁴¹Dyck *et al.* (2007) find that, in the context of U.S. corporate fraud, access to information and monetary rewards play an important role in encouraging whistle-blowing.

no pre-registration is needed since one just needs to have a SSN; but enrolling online might be more costly. Over 40 million people asked for SSN receipts more than once.⁴² Over U.S. \$1.1 billion has been distributed in tax rebates and lottery prizes. During the period of analysis there was a total of 1,151,518 complaints sent to the tax authority by 135,102 different consumers regarding 134,054 different establishments to the tax authority.

2.3 Data and Sample Definition

In this Section, I briefly describe each data source, and the summary statistics of the data. First, I present the firm-level data and the main outcomes I examine in Section 3. Second, I explain the datasets at the consumer level, and the key variables I use in Section 4. In both cases, I focus on features of the data most relevant for my empirical analysis. Additional details on variable definitions and sample choices can be found in the Online Appendix B.

2.3.1 Firm Data

I use administrative data on de-identified establishment-level tax returns and registry information from the Department of Finance of the state of Sao Paulo, Brazil from January 2004 to December 2011.⁴³ In the empirical for analysis, I restrict attention to the 605,994 firms that submitted tax returns between January 2004 and December 2011, and are registered in a retail or wholesale sector. Also, I aggregate establishments by firms as it is likely that firms maximize their tax planning in a consolidated manner.⁴⁴

Firm characteristics. From the registry of firms of Sao Paulo, the main variable I use is the sector of activity. Sectors are defined according to a 7-digit code of the Brazilian National Classification of Economic Activity (CNAE version 2.1). The retail sectors are all the sectors that start with 47 plus motor vehicle retail under sectors that start with 45 and food services (bars and restaurants) in sectors that start with 56. Wholesale is defined by all sectors that start with 46, plus motor vehicle wholesale under sectors that start with 45. The sector definition is very detailed; for instance, 472 is Retail food, beverages, tobacco; 4722-9 is Retail meat and fish; and 4722-9/01 is Retail meat (butchery). Throughout the paper, *sector* refers to the 7-digit definition.

In the empirical analysis I aggregate the firm-level outcomes by sectors of activity between January 2004 and December 2011. There are 212 sectors: 92 in retail and 120 in wholesale. The sector panel has 20,352 monthly observations.

⁴²Since any SSN holder in Brazil is eligible for the rewards, people in neighboring states may also participate (the total population of Sao Paulo is 42 million). Over 500,000 consumers with online accounts are from municipalities outside the state of Sao Paulo.

⁴³Due to confidentiality reasons, I do not have access to data on audits rate or any information uncovered from audits. In addition, the data were de-identified, and no establishment data were provided from sectors that have fewer than five establishments, or from sectors in which one establishment is responsible for over 90% of the sector's tax revenue of that sector. In the groups of sectors I analyze – retail and wholesale – 126 establishments were excluded from a total of 1,035,933 establishments registered in Sao Paulo over the period of analysis.

⁴⁴All the results are robust to using establishments as the unit of analysis instead. 94.08% of firms in Sao Paulo are a single establishment and the average number of establishments per firm is 1.18.

Reported revenue. The NFP program aims to ensure that firms accurately report final sales. Accordingly, the gross revenue reported by a firm is the key variable directly affected by NFP. In addition, all firms must report their gross revenue to the tax authority on a monthly basis. Other outcomes such as tax liabilities and reported inputs require sample restrictions as described below. Therefore, this variable is the primary outcome in my empirical analysis of firm compliance. For more details on the specific forms used to construct this variable, see the Online Appendix B. Table 1 describes the *firm sample*. Statistics for the *firm sample* include the monthly gross reported revenue by firm for the key groups I use in the empirical analysis.

Tax liabilities. It is possible that firms' tax liabilities do not respond to the policy in the same way as reported revenue as it also depends on reported inputs. The data available to this study, however, have some limitations to look at tax liabilities by firms. The variable I observe in the data is the amount of tax a firm is due to remit to the tax authority, not their tax liability. There could be substantial differences between these two quantities. An important driver of this difference is tax withholding within the VAT chain: part of the tax that is due by an firm is withheld and remitted by a upstream or downstream trade partner. This measurement problem introduces mechanical drops and increases in tax liabilities by firms that are difficult to control for as withholding rules are based on products and I do not observe products. Yet, there are some sectors that are less affected by withholding than others. Therefore, in order to look at tax liabilities, I restrict attention to subset of firms (henceforth *tax sample*) that are in sectors with little tax withholding throughout the period of analysis such that the tax due I observe best approximates the tax liabilities of firms.⁴⁵

Reported inputs and Value added. For firms registered in the VAT (as opposed to the turnover tax regime described in Section 2), the tax returns include reported inputs. In order to understand the change in tax liabilities, it is helpful to analyze changes in expenses claimed and the resulting value added. For this analysis I restrict attention to firms that are registered as VAT throughout the period of analysis in order to be able to observe changes in reported expenses. This sample restriction is important as firms that switch in and out of the VAT will have gaps in their reported expenses due to the the tax form they have to file.

Because the impact on tax liabilities, reported inputs, and value added use subsets of firms, I show the effects on reported revenue for these samples to allow for a direct comparison with the overall sample.

Receipts. The micro-data on receipts captures purchases by consumers between January 2009 and December 2011. For these receipts the data include: month and year it was issued, the total amount spent, and an establishment identifier. Therefore, starting in January 2009, I can calculate by firm how many receipts have a SSN on them and the total value of these receipts. The micro-data with receipts details linking individual to firms before January 2009 was not available to this

⁴⁵To identify sectors less affected by withholding I proceed as follows: for firms in the VAT, which submit more detailed tax returns, I obtained an aggregation of the total values of input and output transactions that are in tax codes related to withholding. I aggregate these firms by sector, and calculate how much of total inputs and sales transactions are affected by withholding during the period of analysis. Then I restrict attention to sectors for which neither the input or output transactions affected by withholding represent more than 1% of the total input or output reported by VAT firms in those sectors. Further details on tax liability measurement can be found in the Online Appendix B.

study.

Complaints. 13 million individuals created an online account at the tax authority’s website from October 2007 to December 2011. From these accounts, it is possible to observe the time and quantity of complaints filed against specific establishments. The whistle-blower analysis in Section 4.1 uses data on the timing of when a firm received its first complaint.

In order to reduce the influence of outliers, I winsorize the firms’ outcomes by their 99th percentile value using the monthly micro-data panel, i.e., I replace all values above the 99th percentile of the reported revenue distribution by the 99th cutoff percentile value. Online Appendix A shows results using alternative top coding thresholds.

2.3.2 Consumer Data

Consumer-level datasets are based on de-identified administrative data from NFP receipts and from online account activity at the tax authority’s website.⁴⁶ Importantly, the consumer-level data was generated by the NFP program. Therefore, there is no “pre-NFP” data on receipts, or any other individual characteristic.⁴⁷ Further details on data sources and measurement can be found in the Online Appendix B.

Receipts. As described above, the receipts data file captures purchases for which final consumers asked for SSN receipts between January 2009 and December 2011. The main variables I derive from the receipts dataset are: (i) *number of receipts*: the total number of SSN-identified receipts that a consumer asks per month; (ii) *total expenditures with a SSN*: the total amount of money spent associated with the SSN-identified receipts, aggregated by consumer, per month; In order to reduce the influence of outliers I winsorize the *number of receipts* and *total expenditure in SSN receipts* by their 99th percentile value using monthly micro-data.

Lotteries. From consumers online accounts, it is possible to observe the receipts they ask, their participation in lotteries and the value of cash prizes.⁴⁸ The main variables derived from the online account dataset are: (i) *number of lottery tickets*: the total number of lottery tickets a consumer holds per month; (ii) and *lottery prizes*: the number of lottery prizes and the value of lottery prizes per month. I restrict attention to 24 lottery draws between July 2009 and June 2011, i.e., 6 months before and after the first and last lottery available for this analysis.

Consumer sample. I restrict attention to 5,028,669 consumers who participated in the lotteries between July 2009 and June 2011. This number includes lottery winners and a 10% random sample of non-winners in each lottery draw. The second panel of Table 1 displays the descriptive statistics for these consumers. All the variables are “unconditional”, i.e., the number of lottery tickets,

⁴⁶For confidentiality reasons, no information that may identify individuals was available to this study. A “scrambled” unique identifier was created for each individual SSN, and no information on names or addresses was provided. Also, for a given receipt, the total amount spent is rounded to the nearest integer, and the final data contains no information on prices or products that were purchased.

⁴⁷The state tax authority has no information on individual income tax records or any other federal tax data. Apart from motor vehicle property information, state tax authorities do not usually collect data on individuals.

⁴⁸All data on approximately 90 consumers who won one of the top 3 lottery prizes of over U.S. \$500 were excluded from the datasets available to this study for confidentiality reasons. See Online Appendix B for more details.

prizes and rebates consider the entire period that consumers could be asking for receipts and participating in lotteries, so these variables take value zero when consumers do not ask for receipts or are not registered yet to participate in the lotteries.

3 The effect of third-party information trails on firm compliance

To investigate the degree to which the availability of third-party information trails introduced by consumer rewards for requesting receipts can improve firm compliance, I begin by exploiting the impact of the introduction of the NFP program on revenue reported by firms using a difference-in-differences (DD) research design. I focus here on firms' reported revenue as it is the margin directly affected by the policy.

The identification strategy exploits variation in treatment intensity from the policy change. I compare two downstream sectors affected differently by the consumer monitoring program: retail and wholesale. NFP targets final consumer sales, so firms that sell mostly to final consumers are more affected than firms selling mostly to other firms. To exploit this difference I compare "treated" retail sectors to "control" wholesale sectors. I use a DD design to estimate changes in reported revenue by firms in each group before and after the implementation of the program.

Figure 1 shows descriptive statistics to motivate the comparison between retail and wholesale. Even though the NFP program is targeted at final consumer sales, consumers who purchased directly from wholesalers and manufacturers could enjoy the same reward benefits as in retail purchases. Figure 1a shows the total number of receipts with a consumers SSN in each group of sectors. The receipt-level data available to this study starts in January 2009 only, but it clearly shows a substantial difference in the magnitude of the number of receipts with SSN between the two groups of firms. Figure 1b shows the share of revenue in each sector that is covered by SSN receipts. For retail sectors, this share reaches 40% in 2011 while it is always below 9% in wholesale sectors in the period of analysis. The trends in both figure also indicate that take up of the policy increased overtime, suggesting that the impact of the policy may also be gradual rather than a sharp change at the onset of the policy.

One advantage of the data is that I observe a long time series of pre-NFP observations of reported revenue changes in the sector groups. Thus, I can shed light on whether a key identification assumption in a DD holds: that trends in potential reported revenue changes are parallel for retail and wholesale sectors. Figure 2a displays changes in total raw reported revenue by group of sectors from January 2004 to December 2011. In this figure, each data point is scaled by the average monthly reported revenue before the introduction of the NFP in October 2007 for the group.

In Figure 2a, retail and wholesale reported revenue changes closely trace each other until program implementation. The vertical lines highlight the key moments in the implementation of the program discussed in Section 2.2. Following implementation, change in reported revenue gradually increases in retail sectors, relative to wholesale sectors. Figure A2 in the Online Appendix A shows that the firms in the excluded sectors upstream behaved similarly to wholesale, which is

consistent with the argument that firms that do more business to business transactions should be affected less by this policy. The gradual change in Figure 2a is consistent with the fact that the program was not implemented at once, and consumer participation increased steadily over time. Since the figure displays raw data, there is quite a bit of variation across months of the year due to the seasonality of consumption. In particular, in retail sectors, reported revenue spikes each December, consistent with increased holiday-related consumption.

In order to measure the effect of the program across time, I run a flexible DD specification that includes 17 time dummies for 6-month windows from 2004 - 2011, using October 2007 (the starting point of the program's implementation) as a reference point, and using data aggregated at the 7-digit level in a balanced panel. Each 6-month window, denoted by k , is associated with a dummy variable $Period_t^k$, which equals one if month t falls within window k :⁴⁹

$$\ln R_{st} = \eta_s + \gamma_t + \sum_{k=-8}^8 \beta^k (Treat_s \cdot Period_t^k) + u_{st} \quad (5)$$

where $\ln R_{st}$ is the log of reported revenue in sector s and time t ; η_s are 7-digit sector fixed effects and γ_t are dummies for each month of each year. $Treat_s = 1$ if sector s is a retail sector, and u_{st} is clustered by sector. This specification allows me to show the treatment effect across time, while controlling for finely-defined time and sector effects.

Figure 2b plots the coefficients and the 95% confidence intervals from estimating equation (5) without a constant. The difference between the two groups is relatively constant before NFP. By the time the program is fully implemented the difference in log reported revenue between the two groups begins to grow. This effect, averaged across all post-implementation periods, can be estimated from a standard DD specification:

$$\ln R_{st} = \eta_s + \gamma_t + \beta Treat_s \cdot Post_t + u_{st} \quad (6)$$

where $Post_t = 1$ if $t \geq \text{October 2007}$ and u_{st} is clustered by sector. Figure 2b displays the estimated DD coefficient $\hat{\beta}$ from estimating equation (6). The results suggest that the NFP program induced a positive and significant 21% increase in reported revenue by firms across the 4-year period following implementation. Because I am exploiting differences in the treatment intensity across firms, the estimated effect is likely a lower bound of the program's impact. The control group was also potentially affected by the policy: either directly from sales to final consumers or indirectly from the self-enforcing properties of the VAT.

In addition, I use the firm-level data to test whether the estimated DD effect in equation (6) is similar when controlling for firm i fixed effects. The empirical strategy is analogous to the sector-level analysis:

$$\ln R_{its} = \eta_i + \gamma Post_t + \beta Treat_{ts} \cdot Post_t + \varepsilon_{its} \quad (7)$$

⁴⁹For instance, $Period_t^0 = 1$ if $t \in [\text{Oct.07}, \text{Mar.08}]$, $Period_t^{-1} = 1$ if $t \in [\text{Apr.07}, \text{Sep.07}]$, and $Period_t^1 = 1$ if $t \in [\text{Apr.08}, \text{Sep.08}]$.

The firm-level regression is run in a two-period DD, for which the t is collapsed by *pre* and *post*. The *pre* period is between January 2004 and September 2007, and the *post* period is between October 2007 and December 2011. This strategy avoids log of zero values in firms' monthly data, and helps address serial correlation issues when computing standard errors (Bertrand *et al.* 2004). The regressions are dollar-weighted – i.e., each observation is weighted by its pre-NFP value – such that each observation contributes to all regression estimates according to its economic scale to best approximate the sector aggregate-level analysis. $\ln R_{its}$ is the log of reported revenue where in firm i in period t and sector s . The error ε_{its} is clustered by sector.

Table 2 column [1] shows the DD coefficient for the firm-level regression. It is comparable to the aggregate effect: a 25% increase in reported revenue for retail firms compared to wholesale firms. This increase in compliance is sizable, and shows that incentives to consumers can indeed change firm's ability to under-report sales. The implications of this increase in compliance for tax revenue, however, are not obvious. There are two separate issues: (i) such an increase in reported revenue may not generate a similar increase in tax revenue depending on how reported inputs are adjusted, and (ii) the net increase in taxes can be lower as the government is forgoing tax revenue through rewards. Section 5 discusses these two points in detail.

I conduct a number of robustness checks for both the sector and firm-level regressions reported in the Online Appendix A Table A1. The results are robust to winsorizing the top 5% or the top 0.1% to deal with the influence of outliers, and for clustering standard errors by firm instead of sector in the firm-level estimation. In addition, in Figure A3 I investigate whether the retail-wholesale comparison is indeed capturing an increase in compliance from the NFP policy, rather than an increase in actual revenue or a nation-wide change in trends across the two groups of sectors. Based on aggregate numbers from the tertiary sector annual survey,⁵⁰ the difference between retail and wholesale revenue is constant across time in Brazil, so there does not seem to be a nationwide differential change in revenue between the two groups.⁵¹ Moreover, there is no such differential change in revenue for Sao Paulo firms in the survey data, which suggests that the effect observed in the tax data is indeed a reporting effect and not a real change in economic activity.

4 Mechanisms: whistle-blower threats and collusion costs

In order to investigate the mechanisms through which a consumer reward policy can improve compliance, I turn to the micro data on firms, receipts, and consumers following the predictions from the conceptual framework in Section 1. First, I study the role of whistle-blower threats by examining heterogeneous effects of the program, and by analyzing the behavior of firms after con-

⁵⁰PAC ("Pesquisa anual do comércio") from the Brazilian Census Bureau (IBGE) is an annual national survey conducted by IBGE based on a sample of formal firms in Brazil. The information reported to the survey can more-accurately capture real economic activity as the survey data is highly confidential, and cannot be used to cross check information submitted to the government by firms.

⁵¹The time period post-NFP overlaps with the great recession in the U.S.. The Brazilian economy, however, was not as affected by this particular financial crisis during the period of analysis, and the survey data does not suggest heterogeneous effects across the two groups of sectors.

sumers blow the whistle. Second, I investigate the role of collusion costs. I discuss heterogeneous effects of the policy that could be linked to fixed costs from frictions in setting a collusive deal, and I analyze behavioral biases with respect to reward value that may amplify individual responses to rewards, making it more costly for firms to match the government incentives.

4.1 Whistle-blowers

Heterogeneous effects. I examine heterogeneity in the responses of firms to the NFP policy in order to shed light on how the government can credibly harness the information consumers have on firms' evasion to improve compliance. I begin by allowing the coefficient in specification (7) to be heterogeneous depending on the firm size distribution before the program. The sheer size of a firm could deter under-reporting even absent of incentives to consumers since the number of third-parties firms interact with can have a monitoring effect as discussed in Section 1. Table 2 column [2] shows the DD coefficients separately for firms above (*large firms*) and below (*small firms*) the median firm size as measured by the pre-program reported revenue. The results are in line with the idea that the program affected more small firms that were likely evading more in the baseline.

Then, I use the *number of different consumers* a firm typically faces in their sector of activity to capture the increased detection probability under consumer monitoring: the larger the number of consumers the more likely it may be that one of those consumers will blow the whistle when the firm evades taxes. To construct this measure, I need to use data from the program, i.e., after implementation. In order to avoid using firm-level information that may reflect the treatment effect, I use sector level variation. To define the *number of different consumers* I count the number of different unique SSNs per firm from the receipts data, and I rank retail sectors by the average number of unique SSNs per firm. I take this source of variation to the data using the following specification:

$$\ln R_{its} = \eta_i + \gamma Post_t + \sum_{m=1}^2 \alpha_m (d_{ms} \cdot DD_{ts}) + f(x_i) \cdot DD_{ts} + \varepsilon_{its} \quad (8)$$

As in specification (7), I run this regression as a two-period DD, for which the data is collapsed by *pre* and *post*. Firm fixed effects are denoted by η_i . The term $f(x_i)$ is a 3rd-order polynomial of firm size as measured by the average reported revenue three years before the program, and DD_{ts} variable is the interaction between a dummy for retail sectors and a dummy that equals 1 for post October 2007. The error ε_{its} is clustered by sector. The dummy $d_{1s} = 1$ if sector s is below the median of the *number of different consumers* distribution across sectors, and $d_{2s} = 1$ if sector s is above the median. I flexibly control for firm size effect through an interaction of DD_{ts} with $f(x_i)$ to separate the overall size effect discussed above from the effect of number of different consumers as discussed in the conceptual framework.⁵²

⁵²These results are robust to winsorizing the dependent variable at different cutoffs (0.1% and 5% instead of 1%). Standard errors are also robust to clustering at the firm or time level instead of sector. See Table A2 in Online Appendix A. Similarly to the average effect, the heterogeneous results are assumed to share a common trend with the overall sample of wholesalers conditional on heterogeneous effects by firm size.

Table 2 column [3] shows that the effect is concentrated among firms that face a high *number of different consumers*. This program is changing the availability of information trails, and the threat imposed by potential whistle-blowers might help to explain how this program can work despite collusion opportunities between the buyer and the seller.

Whistle-blower event study. The evidence above indicates that whistle-blower threats could be an important device to improve compliance. In order to further examine the effect of whistle-blowers, I exploit a direct link between the participation of consumers in the enforcement effort and firm behavior. I use a dataset with over 1 million complaints to analyze how firms respond after a consumer blows the whistle. It is worth noting, though, that there does not need to be a link between the number of complaints and the size of the effect of the policy in equilibrium. For instance, in an extreme case that the policy shifts all firms to full compliance, no complaints would be observed in the data.

The degree to which a firm responds to a consumer blowing the whistle depends on their beliefs. If firms believe that the probability ε from Section 1 that a consumer will blow the whistle is too low - perhaps because they have never seen a consumer blow the whistle - they perceive ε as being lower than it actually is. If this is the case, once firms observe consumers blowing the whistle for the first time, they may update upwards their beliefs about ε , which increases compliance. Note that the beliefs do not have to be biased in a specific direction at baseline, just that some firms have beliefs that are too high and some others that are too low. Firms with beliefs that are too high may only learn very slowly (or even never) about the true probability because they will mostly comply. Firms with beliefs that are too low will learn more quickly about the true probability because at the beginning they may not comply as a result and will likely receive complaints. Therefore, it is possible that, on average, firms will update their beliefs upward after a first complaint, and thus increase compliance.

In the data, every month, a firm may receive a complaint from a consumer through the NFP website. A firm is typically notified by a complaint up to one month after it is submitted by the consumer. In order to study the effect of consumers' complaints I examine the impact of the first complaint a firm gets from a consumer through the website. Different firms received their first complaints at different points in time, and I can exploit the timing of the first complaint to assess the response of firms. The likelihood of receiving a complaint in a given point in time, however, may be driven by the volume of sales leading up to the first complaint. Moreover, it is possible that firms that have a large volume of sales in a given month may be followed by a lower reported revenue in the next period due to mean reversion or other seasonal characteristics. Therefore, exploiting the timing of the complaint alone might not be ideal, as subsequent changes in reporting patterns after the first complaint might reflect real changes in economic activity of the firm.

In order to circumvent mean-reversion and other seasonal effects, I build a counterfactual for each complaint event. I create an "event-control" group composed of firms that did not receive their first complaints by a given event date. I.e., these firms may have received complaints after event time zero, but not before or at event time zero. I use a subset of the firm sample defined in

Section 2.3. I consider only retail firms, and within retail I only retain the firms that did not exit before 2009. I use complaints that were filed between July 2009 and June 2011. Throughout this period, 134,054 or 25% of establishments received at least one complaint. Online Appendix B has more details about the samples.

For each firm, I consider the first complaint event as the first time any of its establishments received a complaint. I use a re-weighting method based on quartiles of the propensity score of getting a complaint in a given period to control for firm characteristics and past outcomes. For a detailed description of the propensity score and reweighting see the Online Appendix B.⁵³

Let $i \in \{T, C\}$ index each firm as “complaint” T or a “no-complaint” C in a given month. Let t_o index the month in which an outcome is observed, and t_e index the month in which a consumer blows the whistle on the firm for the first time (the “event-month”). Define $k \equiv t_o - t_e$ as the number of “periods” or months after/before the first complaint. I performed this re-weighting exercise separately for each month between July 2009 and June 2011, and I collapsed the data by event-month $k \in [-6, 6]$ using the propensity score weights for Figure 3.

Figure 3a displays the number of receipts *complaint* and *no-complaint* firms report to the tax authority, and Figure 3b shows changes in reported revenue relative to 6 months before the first consumer blows the whistle. The x-axis shows the distance in months k to the first complaint or “event-month.” The graph also displays the estimated DD coefficient from estimating the following equation in the micro-data:

$$\ln Y_{iekt} = \gamma_{ie} + \pi_t + \phi_k + \beta \cdot I_{iekt} \{k \geq 0, ie = T\} + u_{iekt}, \quad (9)$$

where $\ln Y_{iekt}$ is either the log of the number of receipts or the log of reported revenue that firm i reports to the government in calendar month of event e . The event-month is indexed by k and calendar date is indexed by t . I control for calendar time fixed effects π_t , event-month fixed effects ϕ_k and for firm-event fixed effects γ_{ie} . The same firm i can be in T or C depending on the event as the control group draws from firms that did not yet receive a complaint by event e . Therefore, I can control for firm-event fixed effects to make sure I am only using variation within firm and event. Figure 3 displays the estimated DD coefficient $\hat{\beta}$ from estimating equation (9) for a window $k \in [-6, 6]$ around each event. Standard errors are clustered by event e and are robust to clustering by firm or by firm-event.

I find a significant 7% increase in the number of receipts firms issue and a significant 3% increase in reported revenue after firms receive their first complaint. The impact of the first complaint is capturing the overall impact of receiving complaints, as some firms received additional complaints after time zero. It can be interpreted as an increase in the perceived detection probability as firms learn that consumers can indeed share information with the government about their non-compliance. Audit probabilities could change as a result, but even if audit rates do not change,

⁵³The propensity score of a firm receiving its first complaint at a given time is estimated for each complaint date using age of the firm, number of establishments by firm, dummies for legal nature of the firm, sector fixed effects, dummy for location in the metropolitan region of Sao Paulo, and the three lags of third-order polynomials of reported revenue, reported receipts, SSN receipts and number of consumers.

firms could perceive a higher risk of getting caught as the government is better informed. Note that it should be expected that this whistle-blower effect is smaller than the overall effect in the aggregate analysis because firms have likely already increased compliance prior to any complaint, in anticipation of the risk that consumers will blow the whistle if they do not change compliance at all.

Together, the impact of consumers blowing the whistle and the heterogeneous effect of the consumer-monitoring is consistent with the argument that whistle-blowers can be an important part of the explanation for why third-party reporting is so effective to ensure compliance. In the context of NFP, it can be a tool for the government to tap into the wealth of information that consumers elicit when asking for receipts from hard-to-tax firms that self report final sales.

In the case of NFP, 1% of consumers filed complaints about 20% of firms. This is not necessarily surprising as the number of consumers is much larger than the number of firms, but it highlights how this diffuse monitoring mechanism can improve enforcement even when most consumers are not willing to actively participate in complaints. Arguably firms do not know which consumers among many that are asking for receipts are willing to be whistle-blowers, so the government can exploit this information asymmetry to generate a deterrence effect from this diffuse monitoring.

4.2 Collusion costs

As discussed in the conceptual framework, collusion can be costly above and beyond the additional risk of getting caught from whistle-blower threats. First, if there are any frictions in setting a collusive deal, there could be a fixed cost per transaction that is being concealed. Second, the more consumers value rewards, the more firms need to compensate consumers in a collusive deal through discounts, which decreases the returns from evasion. In this subsection, I discuss heterogeneous results of the policy on firms that may face different collusion costs. Then, I provide evidence on consumer's responses to the lottery rewards. I show that the program is salient by exploiting variation in the disbursement schedules of the monetary rewards. Then, I exploit variation from the monthly lotteries to investigate whether potential behavioral biases with respect to lotteries may amplify the response consumers have from rewards.

4.2.1 Fixed costs of collusion

The comparative statics discussion of section 1 distinguishes between the number of consumers (that should affect whistle-blower threats as discussed above), and the number of transactions or value of transactions (that should affect the cost of evasion for a given firm size). In order to shed light on this channel, I analyze heterogeneous responses to the policy by *volume of transactions* and the *receipt value*. To construct these measures, I follow the same logic as in the *number of different consumers* described above. To define the *volume of transactions* I rank sectors based on the count of the average number of receipts per firm in retail. To define *receipt value* I calculate the median receipts per firm in retail and then I use the median value by sector to rank retailers.

Table 2 columns [4] and [5] show the results using specification 8 and dummies for sectors above and below the median *volume of transactions* and *receipt value*, respectively. The effect of the program is stronger for firms in sectors with high 'foot-traffic': firms in sectors that have a high volume of transactions and small ticket items are affected more controlling for differential effects by firm size. These results are consistent with collusion costs being part of the mechanism through which third-party information can affect compliance. Although there is room for collusion, the policy can introduce a concealment cost that dis-proportionally affect firms that must collude multiple times to continue evading.

The heterogeneous effects by *number of different consumers*, *volume of transactions* and *receipt value* are picking up similar variation in the data, and it is difficult to identify the relative importance of each channel. Still, these patterns help shed light on the different mechanisms through which this policy can operate. These mechanisms could be related: the concealment costs are arguably due to the secretive nature of collusion that is, arguably, a consequence of whistle-blowers concerns. If firms could openly announce clear discount policies to consumers, these fixed costs would be mitigated but they would run a higher risk of getting caught.

4.2.2 Consumers responses to rewards

Now, I turn to the value of rewards to investigate how consumer behavior may contribute to the impact of the policy.

Are consumers paying attention to the rewards? I verify that the release of monthly lottery results is salient to consumers by examining changes in the volume of Google searches about NFP. Google data aggregates information from millions of searches, and they can meaningfully capture salient social patterns that other survey methods cannot capture as easily (Stephens-Davidowitz 2014).⁵⁴ Around the 15th of each month, the tax authority performs the lottery draws and releases information on lottery winners. A consumer can only check her lottery results by logging in to her online account at the tax authority's website. The actual address is not straightforward to remember (<http://www.nfp.fazenda.sp.gov.br>); as a result, consumers looking for this address may search for the program's name or initials.

Figure 4a pools Google search data from the first to the last day of each month between 2008-2011, and it scales each data point by the first day of the month.⁵⁵ From the figure, it is clear that there is an increase in search volume around the 15th of the month the tax authority releases the results of the lotteries: it is 16% higher than on the first day of the month. The gray line displays data from searches with the word "futebol" (soccer in Portuguese) which provides a metric of how the general volume of Google searches varies within a month. Figure 4b shows that the timing of disbursement is also salient: the total amount of rewards requested for bank account deposits

⁵⁴Hoopes *et al.* (2015) use Google and Wikipedia searches about U.S. income tax to show that the propensity to search varies systematically with tax salience.

⁵⁵I exclude the months of April and October – during which the government disburses the tax rebates – to make sure that the search pattern is related to the lotteries.

spike as soon as tax rebates become available every April and October.⁵⁶

The lottery effect. I exploit variation from the monthly lotteries to investigate whether potential behavioral biases with respect to lotteries may amplify the response consumers have from rewards. As detailed in Section 1, the more consumers value the rewards $\kappa(\alpha)$, the more effective NFP will be in preventing tax evasion for a given reward α . The lottery component of the rewards may leverage consumers' taste for gambling or individual behavioral biases. [Friedman & Savage \(1948\)](#) noted that many governments consider lotteries an effective way to raise revenue as individuals may be willing to pay for lotteries paying a negative expected value. [Filiz-Ozbay et al. \(2015\)](#) find evidence that prize-linked savings offered by commercial banks and governments around the world may be more effective at increasing savings than regular interest payments with the same expected value.

In addition, the NFP monthly lotteries typically have three very large prizes – the top prize can be as large as U.S. \$500,000 – and millions of small prizes, which is a payoff structure commonly seen in gambling games and prize-linked savings accounts ([Guillén & Tschoegl 2002](#)). The skewness of the prize values may be a tool to create salience. [Bordalo et al. \(2013\)](#) argue that when comparing alternative risky lotteries, individuals pay attention to the payoffs that are most different relative to their objective probabilities. If consumers exhibit behavioral biases with respect to the NFP lotteries, it would be more difficult for firms to try and replicate the government's rewards to avoid truthfully reporting their sales.

In order to shed light on the role of behavioral effects I exploit the random variation in lottery wins to document consumer participation responses to lottery rewards. Consumers may use past wins as a signal of their likelihood of getting a lottery prize, which would be consistent with misperception of randomness and the use of heuristics in making choices under uncertainty. [Guryan & Kearney \(2008\)](#) find that consumers increased their estimate of the probability that a ticket bought from the store that sold a winning ticket in the past would be a lottery winner (the "lucky store effect").⁵⁷ I restrict attention to small cash prize wins of 5 dollars to investigate whether consumer participation increases after a lottery win. Because the value of the prize is small, a systematic change in behavior after a lottery win is arguably due to an increase in the perceived returns to participation in the program and not an income effect.

To analyze the effect of lotteries I create a natural "event-control" group composed of people that held the same number of lottery tickets in a given lottery but did not win prizes. I use the *consumer sample* defined in subsection 2.3.2. Let $i \in \{T, C\}$ index each consumer as "winners" T or a "non-winners" C in a given month. I use a re-weighting method based on [DiNardo et al. \(1996\)](#) to flexibly control for the number of lottery tickets individuals hold. I create bins for each possible number of lottery ticket holdings up to 40 tickets, which is the set of lottery tickets for which there

⁵⁶As described in the previous section, the tax authority disburses tax rebates biannually. Consumers can use rewards in other ways – e.g., they can be transferred to a third party, used to pay other taxes or saved for a later deposit – so the total amount in the graph will not necessarily add up to the total amount available to consumers at that point in time.

⁵⁷They argue that consumers may rationalize the observed streaks by inferring heterogeneity in the data generating process. In the context of financial investments, [Kaustia & Knüpfer \(2008\)](#) find evidence of reinforcement learning in investors' behavior: personally experienced outcomes are overweighted in future choices.

is common support between the two groups.⁵⁸ I then re-weight the non-winners group such that each bin carries the same relative weight as the analogous bin in the winner group distribution across lottery ticket holdings. This method ensures that I use the random component of the lottery by matching the two groups based on the odds of winning prizes.⁵⁹

Let t_o index the month in which an outcome is observed, and t_e index the month in which the consumer wins the lottery (the “event-month”). Define $k \equiv t_o - t_e$ as the number of “periods” or months after/before the lottery win. I performed this re-weighting exercise separately for each of the 24 lotteries of 5 dollar prizes. I then collapsed the data for each lottery by group and period $k \in [-3, 6]$ using the DFL weights for Figure 5.

Figure 5a displays the average number of receipts for which lottery winners and non-winners ask before and after winning a 5-dollar prize. The x-axis shows the distance in months to the lottery k . Each graph displays the estimated DD coefficient from estimating the following equation in the micro-data for $k \in [-3, 6]$:

$$y_{jekt} = \gamma_{je} + \pi_t + \phi_k + \beta \cdot I_{jekt} \{k \geq 0, je = Win\} + u_{jekt}, \quad (10)$$

where y_{jekt} is the number of SSN receipts or the total value of receipts consumer j asks in “event-month” k and calendar month t . I control for calendar time fixed effects π_t , event-month fixed effects ϕ_k and for consumer-lottery draw fixed effects γ_{je} .⁶⁰ Standard errors are clustered by lottery draw, and are very similar if clustered by consumer.

Figure 5a shows that there is a significant difference in consumer participation in the program between lottery winners and non-winners as measured by the number of receipts they ask with their SSN. The evidence is consistent with a behavioral explanation, given that there is a significant 0.07 difference (.5% increase) between the number of receipts lottery winners and non-winners ask after winning a U.S. \$5 prize, and effect persists after at least 6 months. Since the odds of winning are independent of past wins, the change in behavior observed in Figure 5a suggests that lottery wins could be working as a nudge by making the odds of winning more salient and reinforcing the propensity to ask for receipts. Alternatively, consumers could be using the past lottery win as a signal of luck, and therefore perceive a higher expected return from participating in the program.⁶¹

Figure 5b shows the effect of a 5-dollar win on the total value of receipts. There is a persistent increase of U.S.\$2.5 dollars on average in monthly expenditures after the lottery win. This increase

⁵⁸Figure A5 in the Online Appendix A shows an example of the distribution of lottery ticket holdings among winners and non-winners. It is clear that the winner group typically holds more lottery tickets. Since the number of lottery tickets is determined by consumers’ participation, it is important to carefully control for the odds of winning.

⁵⁹For a detailed description of a similar application of DFL-reweighting see Yagan (2015). For more details on re-weighting see Online Appendix B.

⁶⁰The same consumer j can be a winner and a non-winner depending on the lottery draw as the control group is composed by a 10% random sample of consumers that did not win a prize in lottery draw e , but could have won in another lottery draw. Therefore, I control for consumer-lottery draw fixed effects. See Online Appendix B for more details on the sampling and re-weighting.

⁶¹Appendix A Figure A7 shows the same picture for all prize levels. As the size of the lottery win grows, the estimated effect is larger. This pattern indicates that the change in behavior is indeed due to the lottery win. The effect, however, is confounded with the fact that larger prizes are more relevant cash shocks that can increase the level of overall consumption.

is a change in behavior that lasts for at least 6 months after the lottery draw. If I run specification (10) in a collapsed data to observe the total expenditure before vs. after the lottery win, there is a statistically significant increase in SSN receipts value of \$16.14 (s.e. 3.08) dollars after winning U.S.\$5. Therefore, the effect of the lottery win cannot be attributed to the cash prize alone.

An alternative explanation is that consumers use lottery wins as evidence that the program works as advertised. In the Online Appendix A I study the effect of a U.S. \$5 win for a sample of individuals that won the lottery once before, in which case the effect of confirming that the program works should not be as relevant. Another alternative explanation is that individuals that do not win the lottery get discouraged from the lottery loss. To try to control for this issue, I compare winners of \$5 prizes to \$10 prizes. The difference between the two groups is still a 5-dollar cash prize, but both won the lottery. In both comparisons I find a statistically significant difference in the number of receipts consumers ask after the lottery win that represents a 0.4% change, which provides further support for the interpretation that a small cash prize can change winner's behavior through a higher perceived returns to participation in the program. The results are reported in Figure A6a and A6b in the Online Appendix.

The data does not allow to tease out the exact behavioral bias that the government is exploiting, but the evidence suggests that the lottery component can be a relevant mechanism to explain how NFP can generate enough consumer participation to improve enforcement. Lotteries are used in other contexts such as lottery-linked savings accounts offered by commercial banks, possibility exploiting similar biases. Also, if people misperceive probabilities or simply have a taste for playing lotteries, it would be more costly for firms to match government's incentives. Not only it could increase the necessary discount to make consumers let go of the receipts, it could also create a friction in pinning down the right discount level that may contribute to the collusion costs.

The relative effectiveness of lotteries compared to tax rebates would be a relevant comparison for a cost-benefit analysis, but the variation in the data does not allow to distinguish the two in a compelling manner. As discussed in Section 1, tax rebates could also be leveraging a behavioral effect if framing an additional gain in a separate category ("rebate") is valued more than an cash equivalent discount. Evaluating the relative effectiveness of different reward systems is an interesting avenue for future research.

5 Implications for tax policy

The results in the previous sections show that incentives for consumers to ensure that firms accurately report transactions can be an effective way to improve firm compliance in final sales transactions. The implications for tax policy and its welfare consequences, however, require additional analysis. The effect on tax revenue does not necessarily have to mirror the increase in reported revenue as it depends on the extent to which expenses can be adjusted. Moreover, even if the effect of the policy on tax revenue is positive, it is crucial to evaluate such increase net of consumer rewards. There are also a number of additional costs and benefits for the government, firms and consumers

that should be considered in a welfare analysis. I begin this section by investigating the impact of the policy on tax revenue. Then, I discuss the welfare implications of different components of consumer rewards programs.

5.1 Tax revenue implications

It is possible that firms' tax liabilities do not respond to the policy in the same way as reported revenue. In fact, the response could be proportionally larger or smaller. It could be larger if, for instance, firms in the VAT do not adjust their expenses. In this case, the value added would increase proportionally more than the reported sales. The effect could also be smaller if expenses are adjusted to offset the increase in reported sales.⁶²

In the context of enforcement of sales reporting in corporate income tax, Carrillo *et al.* (2017) and Slemrod *et al.* (2017) find that reported costs increase as well, partially offsetting the change from more accurate revenue reporting. In both cases, the cost increase occurred primarily in difficult to verify margins such as "Other expenses". In the case of the VAT, the ability to adjust inputs tax credit is arguably relatively more limited as a tax credit must be another firms' tax debit. However, VAT credit fraud is often a problem in VAT systems (Bird & Gendron 2007), and there could be under-reporting of inputs if buyers colluded with suppliers to mis-report transactions (Pomeranz 2015).

In order to investigate the effect of the policy on tax revenue, I begin by looking directly at the effect of the policy on firms' tax liabilities. I focus on a subset of sectors with little tax withholding such that the total tax due reported by firms is a good measure of the total tax liability of a firm (*tax sample*).⁶³ First, I run the same analysis as in Section 3 using the flexible DD specification (5) in the sector panel and the log of tax liabilities as an outcome. Figure 6a shows the difference in tax liabilities between retail and wholesale sector. Similarly to reported revenue in Figure 2b, there are parallel pre-trends before the introduction of the policy, and a clear increase in tax liabilities in retail relatively to wholesale after the policy. The DD estimate is a statistically significant 25.9% increase in tax liabilities, which is close to the figure for reported revenue in Section 3.⁶⁴

Table 3 Panel A shows the results using firm-level data and running the DD specification (7) for the tax sample. Column [1] shows the DD coefficient for the log of reported revenue for this subset of firms. The effect is a bit larger than in the main sample, but confidence intervals overlap. Column [2] shows the DD coefficient using the log of tax liabilities, and the results indicate a statistically significant 31.6% increase in reported tax liabilities. Because the liability is zero in some cases, I also use a binary outcome for positive tax liability, but I find no effect on the extensive margin.

⁶²To illustrate this point let value added be $VA = Y - E$, where Y is reported revenue and E is reported expenses, and let $\delta^x = \frac{\Delta x}{x}$ be the change in variable x . $\delta^{VA} = \frac{\delta^Y Y - \delta^E E}{Y - E}$. If $\delta^Y = \delta^E = \bar{\delta}$, $\delta^{VA} = \bar{\delta}$. Also, if $\delta^Y < \delta^E$, $\delta^{VA} < \delta^Y$.

⁶³The data available to this study has some measurement challenges discussed in Section 2.3 and Online Appendix B.

⁶⁴Figure A4 of the Online Appendix A shows the results for reported revenue for the same subsample of sectors used in the tax liability results (*tax sample*). The point estimate is slightly larger (28%) than in the main sample, but the confidence intervals overlap.

The tax liability analysis above is limited to a subsample of sectors, so in order to shed light on the effect of the policy on total tax revenue, I look at changes in the tax revenue in Sao Paulo as a share of GDP compared to all other states combined (leaving Sao Paulo out) in Figure 6b using data from the Brazilian Central Bank. The figure shows a slight level shift in tax/GDP in Sao Paulo of 3.5% relatively to the rest of the country after 2007. This increase is consistent with an effect of the policy on retail tax revenue similar to the 25.9% increase estimated above as taxes in retail are less than 15% of the total tax revenue.

The evidence from both the sector-level analysis and the firm-level data suggest that the percentage change in tax liabilities is similar to the percentage change found in reported revenue. As discussed above, the similarity in the effect is not obvious ex-ante, since it depends on how input claims can be adjusted. Given the evidence so far, any change in reported inputs is not completely off-setting the increase in reported revenue generated by the policy.⁶⁵ In order to investigate reported expenses, I use data from a subset of firms that were in the VAT system throughout the period of analysis.⁶⁶

The Panel B of Table 3 shows the effect of the policy for firms in this subsample using specification (7). In column [1], I show the DD coefficient on the log of reported revenue. The point estimate is also positive and statistically significant. It is again a bit larger than in the main sample but confidence intervals overlap. The DD effect on log of reported inputs in column [2] is significant and only slightly smaller than the effect on reported revenue. The effect on the value added defined by the difference between revenue and input is also significant and similar to the effect on reported revenue. Since some firms have non-positive value added, I also look at a binary outcome for positive value added in column [4]. The effect is not statistically significant. Hence, even though firms do adjust inputs, there is still an increase in the value added, which is in line with the findings above where percentage change in tax liabilities is similar to the percentage change in reported revenue.

The results above suggest that the increase in compliance generated an increase in the effective tax rate. In the Online Appendix D I study real responses to the policy by analyzing formal employment and number of firms in the market by 7-digit sector. The evidence indicates that the increase in tax enforcement did not affect these outcomes during the period of analysis. The null effect may indicate that the implied increase in the effective tax rate is not large enough to affect firms along these margins, and may just reduce evasion rents. The lack of real responses is consistent with the increase in reported revenue being a reporting effect, rather than an actual increase in sales, in which case I could potentially observe an increase in employment or in the number of firms.

⁶⁵For firms outside the VAT - in the turnover regime - an increase in reported revenue would lead to a proportional increase in tax liabilities. However, the majority of the tax collected in retail comes from VAT firms (over 85%), so the adjustment of expenses is still a relevant margin of response.

⁶⁶Firms may switch in and out of the VAT over the period of analysis. When firms are in the simplified tax regime, they do not report inputs as their tax base is turnover. Therefore, it is important to restrict attention to firms that never changed tax regimes to make sure reported inputs can be measured across time. For more details, see Online Appendix B.

Tax revenue net of rewards. The government of Sao Paulo is forgoing part of the tax revenue collected at the final consumer stage by paying the consumer rewards: both incremental revenue from the program, and infra-marginal revenue. Therefore, even if the effect of the policy is positive, it is not clear that the program is able to increase revenue net of transfers.

To perform this calculation, consider the 25.9% point estimate change in tax liabilities from the sector level results. The government is rewarding consumers with 33% of the tax collected in final sales transactions: 30% in tax rebate and 3% in lottery prizes. Considering that these rewards will be applied to 40% of the transactions (as shown in Figure 1), the total revenue increase net of rewards would then be 9.3%.⁶⁷

There are also administrative costs on the government side that should be considered, but there is no official estimate of such costs. Even if they are substantial, it is important to note that the NFP is also arguably relying on the fact that some consumers may never collect rewards. As of 2011, 50% of the rewards were not collected. In particular, there are 27 million consumers that asked for SSN receipts but did not enroll online in the first four years of the program, which is the only way one can claim rewards. Considering the unclaimed rewards, the total tax revenue net of rewards would have increased by 17.6%. This fact highlights two relevant aspects of the policy: (i) the total revenue effect net of rewards could be larger as consumers leave money on the table; (ii) there are non-trivial costs for consumers to fully participate in this policy that should be considered in the welfare implications.

5.2 Welfare discussion

So far, I have focused on the effects of consumer reward policies on firms and their impact on tax revenue. The welfare implications of such policies must also consider social costs and benefits. For instance, improving enforcement could help tilt the playing field in retail away from firms that evade taxes toward the most-efficient firms. Also, it is important to consider how such policies affect the government and consumers beyond redistributive effects of transfers from firms to the government (from additional taxes), or transfers from firms to consumers (through discounts in collusive deals).⁶⁸

The data available to this study does not allow to fully investigate the welfare implications of the NFP, so the aim of this section is to highlight some key points that should be considered in the design of consumer reward policies. Countries adopted different policy bundles, with varying generosity in the reward values and technological sophistication. I discuss the costs and bene-

⁶⁷Consider an extreme case of 100% of receipts with a SSN. In this case, as discussed in the comparative statics of section 1, the government would need an increase in tax collection by almost 50% (i.e., $\frac{0.33}{1-0.33}$) to break even because the reward is applied to all tax transaction at final sales. In the data, 40% of the receipts will receive a reward, so the minimum increase in tax collection to break even becomes 15.2% (i.e., $\frac{0.33*0.4}{1-0.33*0.4}$).

⁶⁸On the firm side, I am abstracting from costs imposed by the program because firms are required to send receipts to the government irrespective of the NFP program. There could be additional time costs from having to enter the SSN digits in the receipts, but I am abstracting from those as there are readily available technologies that could eliminate such costs and are already being adopted in the context of consumer rewards. For instance, the NFP created consumer cards with bar codes that could be scanned to speed up the process. Another example is that the most recent receipt technology has a QR code that can be scanned by consumer rewards smart-phone applications.

fits associated with two relevant dimensions of such policies and their welfare implications: (i) information and tax morale campaigns; (ii) monetary rewards.

Information and tax morale campaigns. It is possible that consumers value tax compliance and have some local information on evasion by firms, but are not fully aware of the importance of official receipts nor of how to volunteer information on tax evasion to the government. In this case, an information campaign that fosters tax morale and emphasizes the importance of issuing receipts and how consumer can whistle-blow firms could leverage a utility gain from a “warm glow” of contributing to an anti-tax evasion program. From section 1, if consumers value receipts even absent of monetary rewards ($\kappa(0) > 0$), they would ask for receipts and affect firm behavior through both mechanisms highlighted in section 4: whistle-blower threats and collusion costs as firms’ would have to offer discounts to compensate consumers for the utility they would gain from getting a receipt.

To fix ideas, assume that such a program without monetary rewards is able to improve firm compliance. In this case, the cost imposed on consumers is low as asking for receipts is driven by a utility gain from the “warm glow” effect, and consumer participation is not being subsidized through rewards. For the government, there would be costs associated to running such campaigns and managing complaints from whistle-blowers. In the case of Sao Paulo, there are no official numbers on the IT and personnel costs to run the NFP program, but new information technologies are making it cheaper for governments to invest in such information channels beyond tax enforcement.⁶⁹ Without monetary rewards to consumers, the cost-benefit of the policy would likely be positive for the government as it is harnessing a tax morale motivation of consumers to ask for receipts, and the government is gathering new information for enforcement. It is worth noting, however, that there could be social costs from this additional surveillance by potentially lowering social cohesion and trust.

Monetary rewards. Most policies of this kind offer rewards to consumers - through tax rebates, cash lotteries or in-kind prizes. If rewards are the main driver of the impact of such policies, it implies that there are costs (e.g., time costs) borne by the consumers that prevent them from asking for receipts, and that the rewards are working as a subsidy to increase the number of receipts requested by consumers. In this case, there could be a dead-weight loss (e.g., from changing time allocation of consumers). Further, the fact that these rewards could be leveraging a behavioral effect complicates the welfare assessment. Considering that the behavioral effect amplifies the perceived size of the subsidy ($\kappa(\alpha) > \alpha$), the change in propensity to participate in the policy would be comparable to that of a higher reward rate. If $\kappa(\alpha) > \alpha$ from a utility gain because consumers enjoy playing the lottery, there is no additional cost to consumers. If rewards are leveraging a mis-perception of consumers (e.g., mis-perception of lottery probabilities), the effective gains would be lower than expected, so consumers could actually be worse off.

For the government, monetary rewards paid to consumers reduce the tax revenue it can obtain. Moreover, there are administrative costs associated with managing the rewards. In this case, re-

⁶⁹For example, in Pakistan, the Punjab Citizens Feedback Monitoring Scheme allows citizens to report petty corruption and other public service delivery issues using their mobile phones [Bhatti et al. \(2014\)](#).

wards that leverage non-financial incentives can be particularly attractive. Taste for gambling and misperception of probabilities make lotteries a cost-effective reward. They are indeed widely used across the world in such programs.

In practice, many programs have a combination of information and tax morale campaigns and monetary rewards. There could be relevant complementarities between the two: if monetary rewards can be thought as a temporary nudge to shift social norms, the program could potentially generate a change in consumer's propensity to ask for receipts even if the government eventually discontinues or reduces the rewards. It is an open question, though, whether there are long term impacts of such programs. In addition, there could be concerns of monetary rewards crowding out intrinsic motivations to ask for receipts. In the case of Sao Paulo, consumers can donate their rewards to charities, which may increase utility from altruistic motives and mitigate these concerns.

In future research, it would be important to build more evidence on the relative cost-effectiveness and welfare implications of different reward options – tax rebates, lotteries with in-kind prizes or cash lottery prizes. The composition of rewards is relevant for the costs of the program to the government, and it is likely key for consumer take-up. A related open question is the critical mass level of consumer take up that such programs need to obtain a sizable enforcement effect, and what is the most cost-effective way to achieve it.

6 Conclusion

Access to substantial third-party information trails is widely believed to be critical for modern tax enforcement. This paper has investigated how the availability of third-party information can improve firms' compliance. I exploit administrative data and quasi-experimental variation from a policy that rewards consumers for ensuring that firms accurately report final sales transactions to the government in Sao Paulo, Brazil.

I find that the program increased revenue reported in retail sectors by at least 21% over four years. I examine heterogeneity across firms and consumer responses to rewards to shed light on the mechanisms. I find that the estimated effect is stronger for smaller firms, for sectors with a high number of different consumers, high volume of transactions, and small ticket items. The findings are consistent with the argument that whistle-blower threats and collusion costs could help explain how self-enforcing incentives can be effective to harness third-party information in a context of extensive opportunities for tax evasion. I also provide direct evidence on the enforcement effect triggered by consumers blowing the whistle: firms report 7% more receipts and 3% more revenue after receiving their first complaint.

Furthermore, I show that consumers are finely tuned to the incentives of the program, and I exploit the random component of lottery rewards to investigate the effect of lotteries on consumer engagement with the policy. I find that consumers condition their participation on past lottery wins. Even small prizes generate a significant and steady increase in the number of receipts con-

sumers request, and the total value of receipts. The results are consistent with the possibility that lotteries amplify consumer responses due to behavioral biases, which would make it more costly for firms to try to match government incentives in order to collude with consumers.

Finally, I study the effect of the policy on tax liabilities. I find that tax revenue increased despite a significant adjustment in reported expenses. I calculate that the policy generated an increase in tax revenues of 9.3% net of rewards.

From a policy perspective, this study sheds light on how citizen engagement can be used as a monitoring tool in hard-to-tax sectors with numerous small taxpayers in a participatory program. In the context of VAT systems, the results indicate that incentives to consumers can potentially help address the last-mile problem of the VAT, which is a well-known shortcoming of one of the most important and prevalent tax instruments in the world.

References

- ALLINGHAM, G. MICHAEL, & SANDMO, AGNAR. 1972. Income Tax Evasion: A Theoretical Analysis. *Journal of Public Economics*, **1**, 323–338.
- ANDREONI, JAMES, ERARD, BRIAN, & FEINSTEIN, JONATHAN. 1998. Tax compliance. *Journal of economic literature*, **36**(2), 818–860.
- ARBEX, MARCELO, & MATTOS, ENLINSON. 2014. Optimal sales tax rebates and tax enforcement consumers. *Oxford Economic Papers*, **67**(2), 479–493.
- BARROSO, RAFAEL, & CORTEZ, JOÃO PEDRO. 2007. Devolvendo impostos para arrecadar mais : uma avaliação dos efeitos da Nota Fiscal Paulista na arrecadação do ICMS no Estado de São Paulo . *Working Paper*.
- BATRA, GEETA, KAUFMANN, DANIEL, & STONE, ANDREW HW. 2003. *Investment climate around the world: Voices of the firms from the World Business Environment Survey*. The World Bank.
- BERTRAND, MARIANNE, DUFLO, ESTHER, & MULLAINATHAN, SENDHIL. 2004. How much should we trust differences-in-differences estimates? *The Quarterly journal of economics*, **119**(1), 249–275.
- BESLEY, TIMOTHY, & PERSSON, TORSTEN. 2013. Taxation and Development. *Pages 51–110 of: AUERBACH, ALAN J., CHETTY, RAJ, FELDSTEIN, MARTIN, & SAEZ, EMMANUEL (eds), Handbook of Public Economics*, vol. 5. Elsevier B.V.
- 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.

- BHATTI, ZUBAIR K, KUSEK, JODY ZALL, & VERHEIJEN, TONY. 2014. *Logged On: Smart Government Solutions from South Asia*. The World Bank.
- BIRD, RICHARD M. 1992. Tax Reform in Latin America: A Review of Some Recent Experiences. *Latin American Studies Association*, **27**(1), 7–36.
- BIRD, RICHARD M, & GENDRON, PIERRE-PASCAL. 2007. *The VAT in Developing and Transitional Countries*.
- BORDALO, PEDRO, GENNAIOLI, NICOLA, & SHLEIFER, ANDREI. 2013. Salience and asset prices. *American Economic Review*, **103**(3), 623–28.
- BRUHN, MIRIAM, & MCKENZIE, DAVID. 2014. Entry regulation and the formalization of microenterprises in developing countries. *The World Bank Research Observer*, **29**(2), 186–201.
- 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.
- COWELL, FRANK. 2004. Carrots and sticks in enforcement.
- DE PAULA, AUREO, & SCHEINKMAN, JOSE A. 2010. Value-added taxes, chain effects, and informality. *American Economic Journal: Macroeconomics*, **2**(4), 195–221.
- DINARDO, JOHN, FORTIN, NICOLE M, & LEMIEUX, THOMAS. 1996. Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach. *Econometrica*, **64**(5), 1001–1044.
- DO AMARAL, GILBERTO LUIZ, OLENIKE, JOÃO ELOI, DO AMARAL, LETÍCIA MARY F., & STEINBRUCH, FERNANDO. 2010. Carga Tributária Brasileira 2009 e Revisão Dos Períodos Anteriores.
- DYCK, ALEXANDER, MORSE, ADAIR, & ZINGALES, LUIGI. 2007. Who blows the whistle on corporate fraud?
- EISSA, NADA, ZEITLIN, ANDREW, KARPE, SAAHIL, & MURRAY, SALLY. 2014. *Incidence and Impact of Electronic Billing Machines for VAT in Rwanda*.
- FABBRI, MARCO. 2015. Shaping tax norms through lotteries. *International Review of Law and Economics*, **44**, 8–15.
- FILIZ-OZBAY, EMEL, GURYAN, JONATHAN, HYNDMAN, KYLE, KEARNEY, MELISSA, & OZBAY, ERKUT Y. 2015. Do lottery payments induce savings behavior? Evidence from the lab. *Journal of Public Economics*, **126**, 1–24.
- FOOKEN, JONAS, HEMMELGARN, THOMAS, HERRMANN, BENEDIKT, *et al.* . 2015. Improving VAT compliance—random awards for tax compliance.

- FRIEDMAN, MILTON, & SAVAGE, LEONARD J. 1948. The utility analysis of choices involving risk. *Journal of political Economy*, **56**(4), 279–304.
- FUND, INTERNATIONAL MONETARY. 2011. *Revenue Mobilization in Developing Countries*. Fiscal Affairs Department.
- 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.
- GUILLÉN, MAURO F, & TSCHOEGL, ADRIAN E. 2002. Banking on gambling: Banks and lottery-linked deposit accounts. *Journal of Financial Services Research*, **21**(3), 219–231.
- GURYAN, JONATHAN, & KEARNEY, MELISSA S. 2008. Gambling at lucky stores: Empirical evidence from state lottery sales. *American Economic Review*, **98**(1), 458–73.
- HADDAD, EDUARDO A, LUQUE, CARLOS A, LIMA, GILBERTO T, SAKURAI, SERGIO N, COSTA, SILVIO M, *et al.* . 2011. Impact assessment of interregional government transfers in Brazil: an inputoutput approach. *Texto para Discussão*, **9**, 2011.
- HOOPEs, JEFFREY L, RECK, DANIEL H, & SLEMROD, JOEL. 2015. Taxpayer search for information: Implications for rational attention. *American Economic Journal: Economic Policy*, **7**(3), 177–208.
- JENSEN, ANDERS. 2016. Employment Structure and the Rise of the Modern Tax system. *Job market paper*.
- JORNAL ESTADO, DE SÃO PAULO. Sep.2007. Sonegação de impostos equivale a 30% do PIB.
- KAUSTIA, MARKKU, & KNÜPFER, SAMULI. 2008. Do investors overweight personal experience? Evidence from IPO subscriptions. *The Journal of Finance*, **63**(6), 2679–2702.
- KEARNEY, MELISSA SCHETTINI, TUFANO, PETER, GURYAN, JONATHAN, & HURST, ERIK. 2010. Making savers winners: An overview of prize-linked savings products.
- KEEN, MICHAEL, & LOCKWOOD, BEN. 2010. The value added tax: Its causes and consequences. *Journal of Development Economics*, **92**(2), 138–151.
- KEEN, MICHAEL, & MINTZ, JACK. 2004. The optimal threshold for a value-added tax. *Journal of Public Economics*, **88**(3-4), 559–576.
- KLEVEN, HENRIK JACOBSEN. 2014. How can Scandinavians tax so much? *Journal of Economic Perspectives*, **28**(4), 77–98.
- KLEVEN, HENRIK JACOBSEN, 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*, **96**(2), 130–134.
- KUMLER, TODD, VERHOOGEN, ERIC, & FRIAS, JUDITH. 2012. Enlisting workers in monitoring firms: Payroll tax compliance in Mexico. *Columbia University Department of Economics Discussion Papers*, **96**.
- LA PORTA, RAFAEL, & SHLEIFER, ANDREI. 2014. Informality and development. *Journal of Economic Perspectives*, **28**(3), 109–26.
- MARCHESE, CARLA. 2009. Rewarding the consumer for curbing the evasion of commodity taxes? *FinanzArchiv: Public Finance Analysis*, **65**(4), 383–402.
- MATTOS, ENLINSON, ROCHA, FABIANA, & TOPORCOV, PATRÍCIA. 2013. Programas de incentivos fiscais são eficazes? Evidência a partir da avaliação do impacto do programa nota fiscal paulista sobre a arrecadação de ICMS. *Revista Brasileira de Economia*, **67**(1), 97–120.
- MONTEIRO, JOANA CM, & ASSUNÇÃO, JULIANO J. 2012. Coming out of the shadows? Estimating the impact of bureaucracy simplification and tax cut on formality in Brazilian microenterprises. *Journal of Development Economics*, **99**(1), 105–115.
- MUSGRAVE, RICHARD ABEL. 1969. Fiscal systems.
- POMERANZ, DINA. 2015. No taxation without information: Deterrence and self-enforcement in the value added tax. *The American Economic Review*, **105**(8), 2539–2569.
- SHLEIFER, ANDREI, & VISHNY, ROBERT W. 1993. Corruption. *The Quarterly Journal of Economics*, **108**(3), 599–617.
- SLEMROD, JOEL. 2007. Cheating ourselves: The economics of tax evasion. *Journal of Economic Perspectives*, **21**(1), 25–48.
- SLEMROD, JOEL, & GILLITZER, CHRISTIAN. 2013. *Tax systems*. MIT Press.
- SLEMROD, JOEL, & YITZHAKI, SHLOMO. 2002. Tax avoidance, evasion, and administration. *Pages 1423–1470 of: Handbook of public economics*, vol. 3. Elsevier.
- SLEMROD, JOEL, COLLINS, BRETT, HOOPES, JEFFREY L, RECK, DANIEL, & SEBASTIANI, MICHAEL. 2017. Does credit-card information reporting improve small-business tax compliance? *Journal of Public Economics*, **149**, 1–19.
- STEPHENS-DAVIDOWITZ, SETH. 2014. The cost of racial animus on a black candidate: Evidence using Google search data. *Journal of Public Economics*, **118**, 26–40.

- THALER, RICHARD H. 1999. Mental accounting matters. *Journal of Behavioral Decision Making*, **12**(3), 183–206.
- TIROLE, JEAN. 1986. Hierarchies and bureaucracies: On the role of collusion in organizations. *JL Econ. & Org.*, **2**, 181.
- TUFANO, PETER. 2008. Saving whilst gambling: An empirical analysis of UK premium bonds. *American Economic Review*, **98**(2), 321–26.
- WAN, JUNMIN. 2010. The incentive to declare taxes and tax revenue: the lottery receipt experiment in China. *Review of Development Economics*, **14**(3), 611–624.
- YAGAN, DANNY. 2015. Capital tax reform and the real economy: The effects of the 2003 dividend tax cut. *American Economic Review*, **105**(12), 3531–63.
- YANIV, GIDEON. 1993. *Collaborated employee-employer tax evasion*. National Insurance Institute, Administration of Research and Planning.

TABLE 1: DESCRIPTIVE STATISTICS

	Number of Obs.	Mean	Std. Dev.	Time Period
Firm sample				
<i>Retail firms</i>				
Reported revenue	37,717,931	18,210	459,832	Jan.2004-Dec.2011
Number of receipts with SSN	12,838,202	174	12,925	Jan.2009-Dec.2011
Number of consumers	12,838,202	106	7,047	Jan.2009-Dec.2011
<i>Wholesale firms</i>				
Reported revenue	3,494,539	95,147	384,970	Jan.2004-Dec.2011
Consumer sample				
Number of receipts	181,032,084	8.75	9.23	Jan. 2009 - Dec.2011
Number of businesses	181,032,084	5.24	5.16	Jan. 2009 - Dec.2011
Tax Rebate	181,032,084	4.66	8.18	Jan. 2009 - Dec.2011
Total expenditure in SSN receipts	181,032,084	502.89	1,060.49	Jan. 2009 - Dec.2011
Number of lottery tickets	181,032,084	1.40	4.86	Jan. 2009 - Dec.2011
Lottery prize value	181,032,084	0.57	3.33	Jan. 2009 - Dec.2011

Note: Tables present descriptive statistics of the main variables for each data source: firms and consumers. All values are in US dollars (US\$1=R\$2) and are measured monthly. *Reported revenue* is the monthly gross reported revenue by firms. *Number of receipts with SSN* is the monthly number of receipts firms report to the tax authority with the Social Security Numbers (SSNs) of the consumer. *Number of consumers* is the total number of different SSN to which firms issue a receipt. The consumer sample includes consumers that participate in at least one lottery between June 2009 and June 2011. *Number of receipts* is the total number of SSN receipts a consumer gets per month. *Number of businesses* is the number of different establishments for which a consumer gets SSN receipts per month. *Tax rebate* is the total tax rebate consumers get from the *SSN receipts*. *Total expenditures in SSN receipts* is the total amount of money (USD) spent with SSN receipts. *Number of lottery tickets* is the total number of lottery tickets a consumer holds per month. *Lottery prize values* is the total amount (USD) of lottery prizes per month. All variables in the consumer sample are assigned a zero when missing for a given time period. For details on the sample and data construction, see Section 2 and Online Appendix B.

TABLE 2: REPORTED REVENUE EFFECT – RETAIL VS. WHOLESALE

	log Reported Revenue				
	[1]	[2]	[3]	[4]	[5]
DD (Post Oct 07 * Retail)	0.254*** [0.0722]				
DD * Large firms		0.253*** [0.0732]			
DD * Small firms		0.350*** [0.0511]			
DD * High volume of different consumers			0.246*** [0.0705]		
DD * Low volume of different consumers			0.0329 [0.0919]		
DD * High volume of transactions				0.253*** [0.0335]	
DD * Low volume of transactions				0.0181 [0.0391]	
DD * High value of transactions					0.0969 [0.0689]
DD * Low value of transactions					0.285*** [0.0754]
3rd-order polynomial of firm size * DD			X	X	X
Time FE	X	X	X	X	X
Firm FE	X	X	X	X	X
Observations	1,035,268	1,035,268	1,035,268	1,035,268	1,035,268
Adjusted R-squared	0.907	0.907	0.908	0.909	0.908

Note: Table 2 displays the main coefficients from firm-regressions. The variable DD is defined as the interaction between a dummy for retail sectors (*Retail*) and a dummy that equals 1 for time periods after Oct 2007 (*Post Oct 07*). The dependent variable is log of reported revenue by firm, and the data is collapsed into two periods: before and after Oct. 2007. Time and firm fixed effects are included in all regressions. The regressions are dollar-weighted (each observation is weighted by the pre-policy reported revenue) such that each observation contributes to all regression estimates according to its economic scale to best approximate the aggregate effect. Column [1] shows the average DD estimate discussed in Section 3. Columns [2] to [5] are discussed in Section 4. Column [2] splits firms in two groups: firms below the median of the baseline firm size distribution (Small firms) and firms above the median (Large firms). Column [3] splits retail sectors into two groups: sectors below the median volume of consumers across sectors (Low volume of different consumers) and sectors above the median of volume of consumers (High volume of different consumers). Volume of different consumers is defined the average number of different SSN reported in receipts by firms in a given sector between 2009 and 2011. Column [4] splits retail sectors into two groups: sectors below the median volume of transactions distribution across sectors (Low volume of receipts) and sectors above the median of volume of transactions (High volume of receipts). Volume of receipts is defined by the average number of transaction by firms in a given sector between 2009 and 2011. Column [5] aims to capture a similar variation as in column [4], but it split retail sectors into two groups based on whether they are in sectors where transaction values are below the median transaction value across sectors (low value of transactions) or above the median transaction value (high value of transactions). Transaction value is defined by the median transaction by firms in a given sector between 2009 and 2011. In order to control for firm size effects the regressions in columns [3] to [5] include a 3rd order polynomial interacted with the DD variable. Size is defined by the average reported revenue by firms during a four-year period before program implementation. Standard errors are clustered at the sector level. See Table A2 in Online Appendix A for robustness checks. Significance levels *** 1%, ** 5%.

TABLE 3: TAX LIABILITY AND REPORTED EXPENSES – RETAIL VS. WHOLESALE

<i>Panel A: Tax sample</i>				
	Log of Reported Revenue [1]	Log of Tax Liability [2]	Positive tax liability [3]	
DD (Post Oct 07 * Retail)	0.311** [0.151]	0.316** [0.137]	0.0434 [0.0350]	
Firm FE	X	X	X	
Time FE	X	X	X	
Observations	167,110	133,950	167,110	
Adjusted R-squared	0.85	0.876	0.801	

<i>Panel B: Expenses, output and value added - firms that were always VAT</i>				
	Log of Reported Revenue [1]	Log of Reported Inputs [2]	Log of Reported Value Added [3]	Positive Value Added [4]
DD (Post Oct 07 * Retail)	0.363*** [0.0824]	0.302*** [0.0833]	0.387*** [0.105]	0.0192 [0.0153]
Firm FE	X	X	X	X
Time FE	X	X	X	X
Observations	88,422	88,422	70,845	88,422
Adjusted R-squared	0.87	0.85	0.90	0.71

Note: Table 3 displays the main coefficients from regressions described in Section 5 using the firm-level data. The variable DD is defined as the interaction between a dummy for retail sectors (*Retail*) and a dummy that equals 1 for time periods after Oct 2007 (*Post Oct 07*). The data is collapsed into two periods: before and after Oct. 2007. Time and firm fixed effects are included in all regressions. The regressions are dollar-weighted (each observation is weighted by the pre-policy reported revenue) such that each observation contributes to all regression estimates according to its economic scale to best approximate the aggregate effect. *Panel A* reports the results for a sample of firms (*tax sample*) that are in sectors where there is little tax withholding and, therefore, the firm-level reported tax liabilities in the data best approximates their own tax liabilities (see Online Appendix B for more details). Column [1] shows the DD results for the log of reported revenue as in Table 2 column [1], but for the *tax sample*. The outcome in column [2] is the log of tax liabilities that excludes non-positive values, and column [3] reports the effect on a binary outcome of whether firms report positive liabilities. *Panel B* reports the coefficients for a sample of firms from all sectors that are registered as VAT throughout the data period (Jan. 2004 – Dec. 2011). Column [1] shows the DD results for the log of reported revenue as in Table 2 column [1] for this subsample. Column [2] shows the DD coefficient for the log reported inputs. The outcome in column [3] is the log of value added that excludes non-positive values, so column [4] reports the effect on a binary outcome of whether firms report positive value added. Standard errors are clustered at the sector level. See Table A3 in Online Appendix A for robustness checks. Significance levels *** 1%, ** 5%.

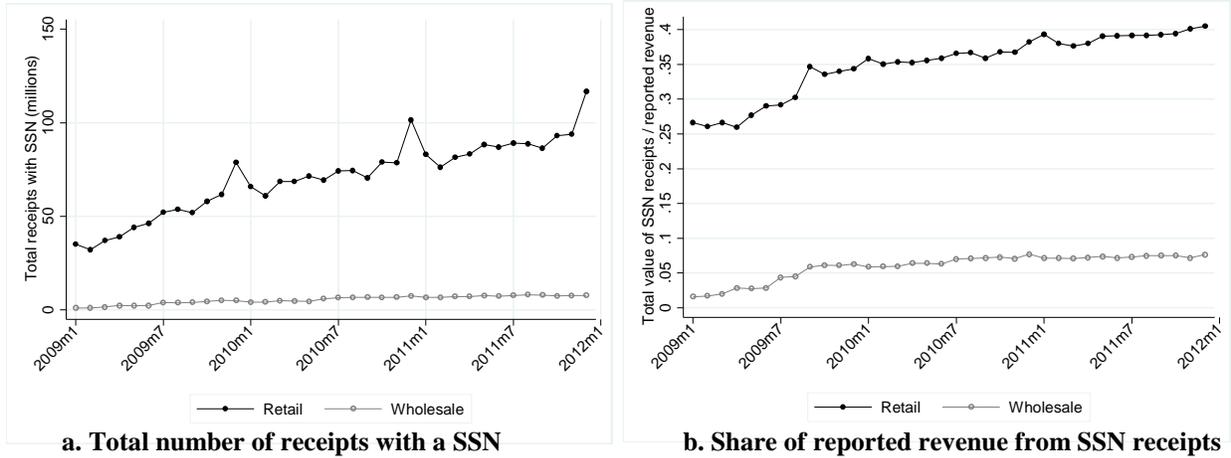


FIGURE 1: NFP RECEIPTS AND REVENUE SHARES – RETAIL VS. WHOLESALE

Note: Figure 1a shows the aggregate number of SSN receipts issued by firms in retail vs wholesale sectors. The figure plots the raw data. The spikes around December of each year follows the seasonal variation in consumption. Figure 1b shows the share of the total reported revenue in retail that is covered by SSN receipts in retail sectors and wholesale sectors. Even though the program was created in Oct.07, the data from the NFP program available for the analysis begins in January 2009. For more details see Section 2 and Online Appendix B. The two figures provide support to the intensity of treatment variation if the difference-in-differences research design discussed in Section 3. They show how Retail sectors were affected relatively more than the wholesale sectors by the NFP program, and that the take up of the program gradually increased over time.

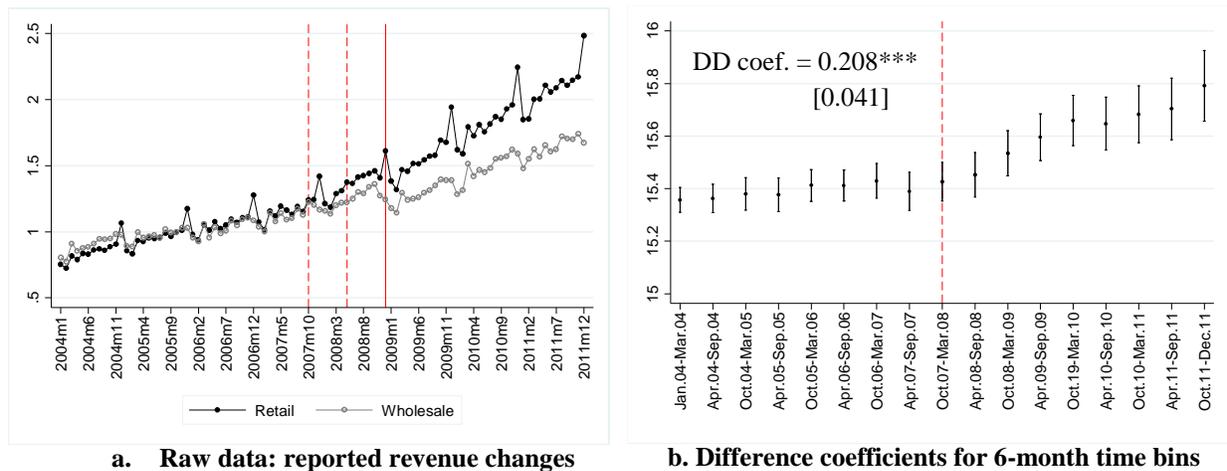


FIGURE 2: EFFECT OF THE POLICY ON REPORTED REVENUE – RETAIL VS. WHOLESALE

Note: Figure 2a shows reported revenue changes for retail and wholesale sectors. Each line is the revenue reported by all firms aggregated by retail or wholesale scaled by the average monthly reported revenue before Oct. 07 for each sector group. The figure plots the raw data. The spikes around December of each year follows the seasonal variation in consumption. The vertical lines highlight the key dates for the implementation of the NFP program: phase-in of sectors begins in Oct.07 and ends in May.08, and the first lottery based on the purchases with SSN receipts was introduced in Dec.2008. Figure 2b plots regression coefficients from estimating specification (5) in Section 3 using a sample of 212 sectors between Jan 2004 and Dec 2011. The sector sample has 20,352 observations. The difference in differences (DD) coefficient displayed in the figure is estimated using the specification (6) in Section 3 where the DD variable is defined by the interaction between a dummy for retail sectors and a dummy that equals 1 for time periods after Oct 2007. Standard errors are clustered by sector. See Table A1 Panel B and Figure A2 in Online Appendix A for robustness checks. Significance levels *** 1%, ** 5%.

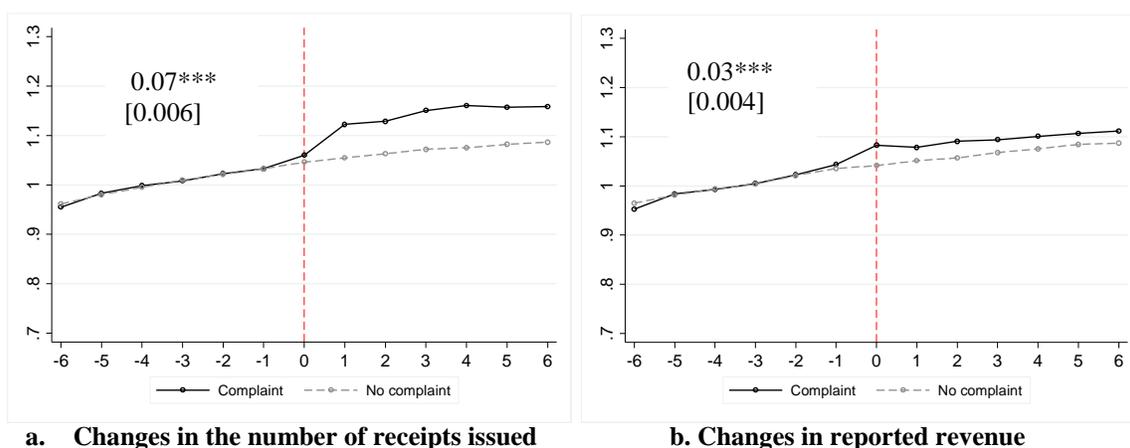


FIGURE 3: WHISTLE-BLOWER EFFECT ON FIRM COMPLIANCE

Note: Figures 3a and 3b plot the changes in the total raw number of receipts firms issue and changes in reported revenue to the government before and after a firm receives the first complaint. Both graphs display changes across event-time where each data point is scaled by the outcome's average before the first complaint (event-time zero). The 'Complaints' group is composed by firms that received their first complaint at event-time zero. The 'No complaint' group is composed by firms that did not receive their first complaint at event-time zero and firms that did not receive a complaint until Dec. 2011. The outcome is averaged across groups and event times using weights based on quartiles of the propensity score to get the first complaint in a given calendar time. The propensity score is estimated using time specific trends for each sector, age of the firm, number of establishments by firm, dummies for legal nature of the firm, sector fixed effects, dummy for location in the metropolitan region of Sao Paulo, and the three lags of third-order polynomials of reported revenue, reported receipts, SSN receipts and number of consumers (see Online Appendix B for more details). The estimated DD coefficient displayed in each graph is based on estimating specification (9) described in Section 4 using the micro-data and clustering the standard errors by the calendar date of the first complaint. Significance levels *** 1%, ** 5%.

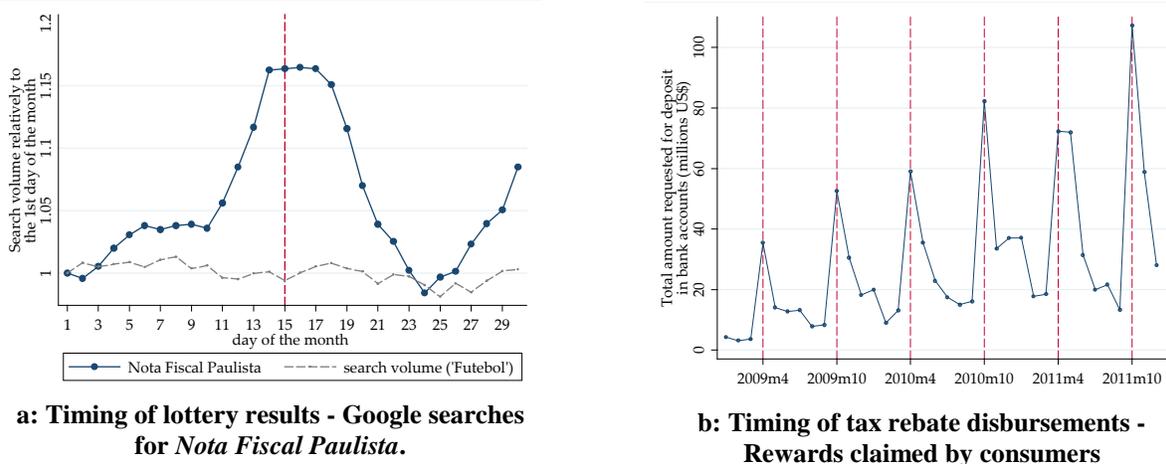


FIGURE 4: ARE CONSUMERS PAYING ATTENTION TO THE REWARDS SCHEDULE?

Note: Figure a displays the search volume from Google Trends website for Google searches with terms related to "nfp" or "nota fiscal paulista" or "nota paulista" pooled by day of the month from IPs addresses in the state of Sao Paulo between 2008 and 2011. It also displays searches for "futebol" (soccer in Portuguese) pooled by day of the month from IPs addresses in the state of Sao Paulo for the same time period. The lottery results are released around the 15th of each month marked by the solid vertical line. As described in Section 2, the tax authority does a biannual disbursement of the tax rebates: every April and October, and creates salience for the program at different dates within the month. Figure b shows the data for rewards claimed across time: each data point is the total amount in millions of US\$ requested for direct deposit in consumer's bank accounts.

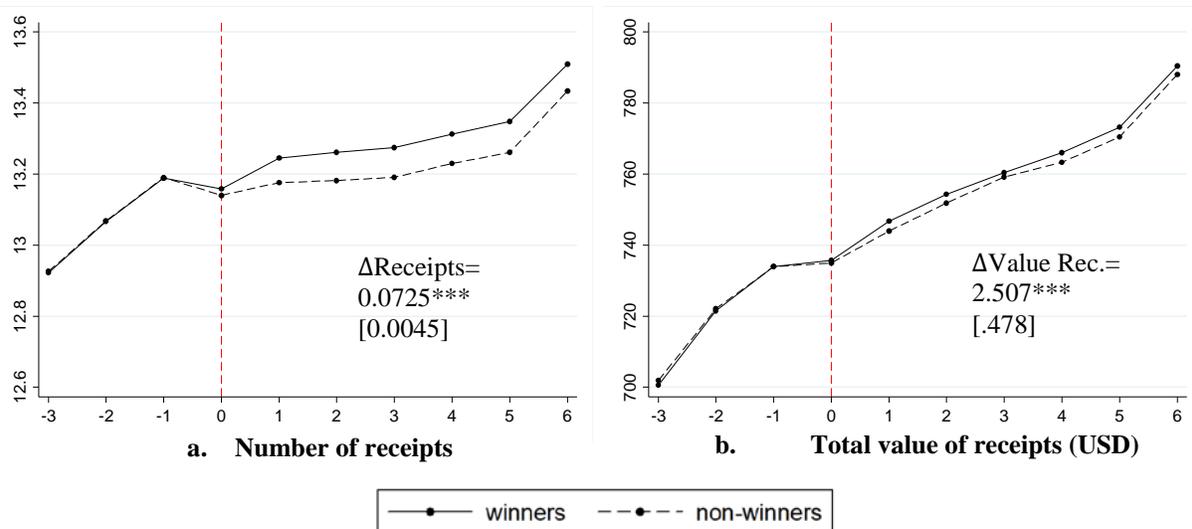


FIGURE 5: THE EFFECT OF A 5-DOLLAR LOTTERY WIN ON CONSUMER PARTICIPATION

Note: The graphs show the raw data by month aggregating all lotteries from June 2009 to June 2011. The x-axis is the number of months since the individual participated in a lottery. The winner group got a cash prize of US \$5 (R\$10) and the non-winner group did not get any prize. Figure 5a plots the total number of receipts consumers ask in each group before and after the lottery draw at event-time zero. Figure 5b shows the total value of receipts (in USD) for each group before and after the lottery draw at event-time zero. Before taking the averages in each case, I create bins for each possible number of lottery ticket holdings from 1-40 tickets in each monthly lottery for 24 lotteries between June 2009 and June 2011. Then I re-weight the non-winners group such that each bin carries the same relative weight as the winner group distribution across lottery ticket holdings (for more details see Online Appendix B). The DD coefficient displayed in each graph is based on estimating specification (10) in Section 2 using the micro-data and the lottery ticket weights. Standard errors are clustered by lottery draw. Significance levels *** 1%, ** 5%.

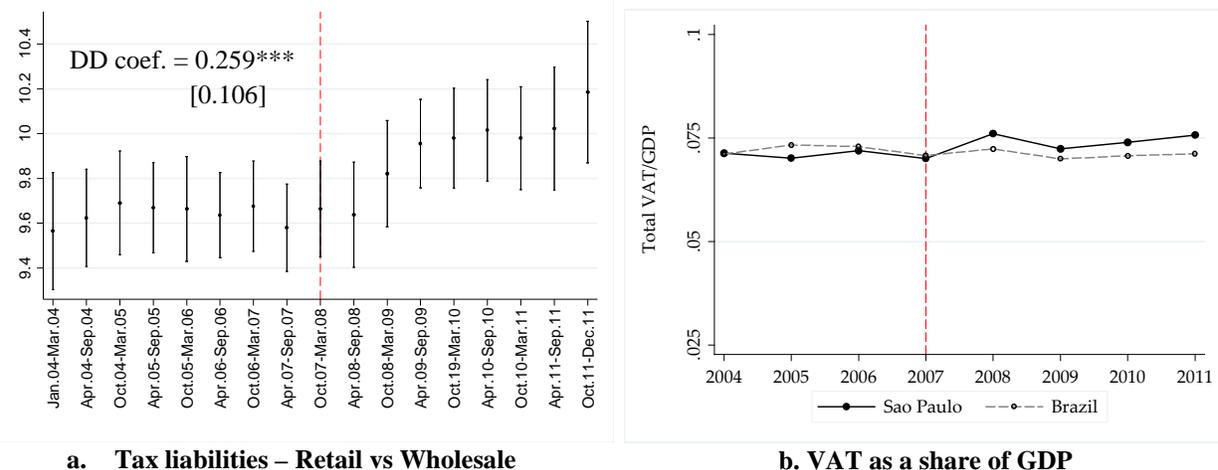


FIGURE 6: EFFECT OF THE POLICY ON TAX REVENUE – RETAIL VS. WHOLESALE

Note: Figure 6a plots regression coefficients from estimating specification (5) using log of tax liabilities as the dependent variable and a sample of sectors for which total tax due best approximates the tax liability of firms between Jan 2004 and Dec 2011 (see Online Appendix B for more detail). Similarly, the difference in differences (DD) coefficient displayed in the figure is estimated using log of tax liabilities as the dependent variable in specification (6). The DD variable is defined by the interaction between a dummy for retail sectors and a dummy that equals 1 for time periods after Oct 2007. This sector sample has 5,088 observations and standard errors are clustered sector. Figure A4 in the Online Appendix A shows the effect of the policy on reported revenue using the same *tax sample*. Figure 6b shows total VAT revenue in Sao Paulo as a share of the state's GDP comparing with total VAT collected in Brazil as a share of the total GDP in Brazil using data from the Brazilian Central Bank. The figures for Brazil include all Brazilian states leaving Sao Paulo out. Table A3 Panel B in the Online Appendix A shows robustness checks. Significance levels *** 1%, ** 5%.

ONLINE APPENDIX A: ADDITIONAL TABLES AND FIGURES

TABLE A1: ROBUSTNESS TOP CODING AND STANDARD ERRORS – MAIN ESTIMATES

<i>Panel A: Log of Reported Revenue - Firm-level regressions</i>			
	p99	p99.9	p95
DD (Post Oct 07 * Retail)	0.254***	0.291***	0.200***
<i>s.e. clustered by sector</i>	[0.0722]	[0.107]	[0.0521]
<i>s.e. clustered by firm</i>	[0.0360]	[0.0690]	[0.0251]
Firm FE	x	x	x
Time FE	x	x	x
Observations	1,035,268	1,035,268	1,035,268
Adjusted R-squared	0.907	0.907	0.88
<i>Panel B: Log of Reported Revenue - Sector-level regressions</i>			
	p99	p99.9	p95
DD (Post Oct 07 * Retail)	0.208***	0.186***	0.249***
<i>s.e. clustered by sector</i>	[0.0411]	[0.0488]	[0.0340]
Sector FE	x	x	x
Time FE	x	x	x
Observations	20,352	20,352	20,352
Adjusted R-squared	0.982	0.976	0.987

Note: *Panel A* displays the robustness of the coefficient reported in Table 2 column [1] from regressions using the firm-level data described in Section 3. The variable DD is defined as the interaction between a dummy for retail sectors (*Retail dummy*) and a dummy that equals 1 for time periods after Oct 2007 (*Post Oct 07*). The dependent variable is log of reported revenue by firm, and the data is collapsed into two periods: before and after Oct. 2007. Time and firm fixed effects are included in all regressions. The regressions are dollar-weighted (each observation is weighted by the pre-policy reported revenue) such that each observation contributes to all regression estimates according to its economic scale to best approximate the aggregate effect. The columns indicate the threshold used to winsorize the dependent variable in order to mitigate the influence of outliers. The results for p99 that are reported in the paper are presented first, then p99.9 and p.95 are also shown. Each column reports standard errors clustered by sector and clustered by firms. *Panel B* displays the robustness of the DD coefficient reported in Figure 2b estimated using the sector-level monthly data. The dependent variable is log of total reported revenue by sector. Time and sector fixed effects are included in all regressions. The columns indicate the threshold used to winsorize the dependent variable in order to mitigate the influence of outliers. The results for p99 that are reported in the paper are presented first, then p99.9 and p.95 are also shown. Standard errors are clustered by sector. Significance levels *** 1%, ** 5%.

TABLE A2: ROBUSTNESS TOP CODING AND STANDARD ERRORS – HETEROGENEITY ANALYSIS

	Log of Reported Revenue									p99	p99.9	p95
	p99	p99.9	p95	p99	p99.9	p95	p99	p99.9	p95			
DD * Large firms	0.253***	0.292***	0.191***									
<i>s.e. clustered by sector</i>	[0.0732]	[0.107]	[0.0524]									
<i>s.e. clustered by firm</i>	[0.0363]	[0.0693]	[0.0254]									
DD * Small firms	0.350***	0.350***	0.470***									
<i>s.e. clustered by sector</i>	[0.0511]	[0.0827]	[0.0453]									
<i>s.e. clustered by firm</i>	[0.0287]	[0.0615]	[0.0225]									
DD * High volume of different consumers				0.246***	0.275**	0.202***						
<i>s.e. clustered by sector</i>				[0.0705]	[0.108]	[0.0508]						
<i>s.e. clustered by firm</i>				[0.0335]	[0.0727]	[0.0242]						
DD * Low volume of different consumers				0.0329	0.0436	0.0712						
<i>s.e. clustered by sector</i>				[0.0919]	[0.115]	[0.0649]						
<i>s.e. clustered by firm</i>				[0.0380]	[0.0701]	[0.0251]						
DD * High volume of transactions							0.253***	0.289***	0.208***			
<i>s.e. clustered by sector</i>							[0.0707]	[0.106]	[0.0516]			
<i>s.e. clustered by firm</i>							[0.0335]	[0.0717]	[0.0241]			
DD * Low volume of transactions							0.0181	0.00226	0.0584			
<i>s.e. clustered by sector</i>							[0.0865]	[0.130]	[0.0614]			
<i>s.e. clustered by firm</i>							[0.0391]	[0.0881]	[0.0268]			
DD * High value of transactions										0.0969	0.153	0.0938*
<i>s.e. clustered by sector</i>										[0.0689]	[0.121]	[0.0531]
<i>s.e. clustered by firm</i>										[0.0424]	[0.0868]	[0.0274]
DD * Low value of transactions										0.285***	0.319***	0.224***
<i>s.e. clustered by sector</i>										[0.0754]	[0.123]	[0.0558]
<i>s.e. clustered by firm</i>										[0.0325]	[0.0655]	[0.0244]
3rd-order polynomial of firm size * DD				x	x	x	x	x	x	x	x	x
Time FE	x	x	x	x	x	x	x	x	x	x	x	x
Firm FE	x	x	x	x	x	x	x	x	x	x	x	x
Observations	1,035,356	1,035,356	1,035,356	1,035,356	1,035,356	1,035,356	1,035,356	1,035,356	1,035,356	1,035,356	1,035,356	1,035,356
Adjusted R-squared	0.907	0.906	0.88	0.908	0.907	0.881	0.908	0.907	0.881	0.908	0.907	0.881

Note: The table displays the robustness of the coefficients columns [2] to [5] in Table 2. The variable DD is defined as the interaction between a dummy for retail sectors (*Retail dummy*) and a dummy that equals 1 for time periods after Oct 2007 (*Post Oct 07*). The dependent variable is log of reported revenue by firm, and the data is collapsed into two periods: before and after Oct. 2007. Time and firm fixed effects are included in all regressions. The regressions are dollar-weighted (each observation is weighted by the pre-policy reported revenue) such that each observation contributes to all regression estimates according to its economic scale to best approximate the aggregate effect. The columns indicate the threshold used to winsorize the dependent variable in order to mitigate the influence of outliers. The results for p99 that are reported in the paper are presented first, then p99.9 and p.95 are also shown. Each column reports standard errors clustered by sector and clustered by firms. Definition of large vs small firms; high vs low volume of different consumers; high vs low volume of transactions; and high vs low value of transactions are the same as in Table 2 and Section 4 of the paper. See notes of Table 2 for more details. Significance levels *** 1%, ** 5%.

TABLE A3: ROBUSTNESS TOP CODING AND STANDARD ERRORS – TAX LIABILITY AND REPORTED EXPENSES

<i>Panel A: Tax sample - Firm-level regressions</i>										
	Log of Reported Revenue			Log of Tax Liability			Positive tax liability			
	p99	p99.9	p95	p99	p99.9	p95				
DD (Post Oct 07 * Retail)	0.311**	0.295**	0.266***	0.316**	0.284**	0.420***	0.0434			
<i>s.e. clustered by sector</i>	[0.151]	[0.139]	[0.0741]	[0.135]	[0.117]	[0.116]	[0.0350]			
<i>s.e. clustered by firm</i>	[0.125]	[0.127]	[0.0518]	[0.137]	[0.149]	[0.0638]	[0.00597]			
Observations	167,110	167,110	167,110	133,950	133,950	133,950	167,110			
Adjusted R-squared	0.851	0.868	0.839	0.88	0.88	0.86	0.801			
<i>Panel B: Tax sample - Sector level regressions</i>										
	Log of Reported Revenue			Log of Tax Liability						
	p99	p99.9	p95	p99	p99.9	p95				
DD (Post Oct 07 * Retail)	0.280***	0.253***	0.282***	0.259**	0.234**	0.319***				
<i>s.e. clustered by sector</i>	[0.0665]	[0.0755]	[0.0502]	[0.106]	[0.106]	[0.104]				
Observations	5,088	5,088	5,088	5,088	5,088	5,088				
Adjusted R-squared	0.983	0.976	0.99	0.956	0.954	0.967				
<i>Panel C: Expenses, output and value added - VAT firms</i>										
	Log of Reported Revenue			Log of Reported Inputs			Log of Reported Value Added			Positive Value Added
	p99	p99.9	p95	p99	p99.9	p95	p99	p99.9	p95	
DD (Post Oct 07 * Retail)	0.363***	0.351**	0.314***	0.302***	0.345**	0.268***	0.387***	0.258**	0.354***	0.0192
<i>s.e. clustered by sector</i>	[0.108]	[0.139]	[0.0931]	[0.107]	[0.151]	[0.0970]	[0.147]	[0.129]	[0.102]	[0.0176]
<i>s.e. clustered by firm</i>	[0.0824]	[0.110]	[0.0540]	[0.0833]	[0.115]	[0.0611]	[0.105]	[0.156]	[0.0592]	[0.0153]
Observations	88,422	88,422	88,422	88,422	88,422	88,422	70,845	70,845	70,845	88,422
Adjusted R-squared	0.87	0.89	0.87	0.85	0.88	0.83	0.90	0.91	0.90	0.71

Note: Table A3 displays the main coefficients from regressions described in Section 5. The variable DD is defined as the interaction between a dummy for retail sectors (*Retail*) and a dummy that equals 1 for time periods after Oct 2007 (*Post Oct 07*). *Panel A* reports the results for a sample of firms that are in sectors where there is little tax withholding and, therefore, the firm-level reported tax liabilities in the data best approximates their own tax liabilities (see Section 2 and Appendix B for more details). The results in *Panel A* show the robustness of the results from Table 3 *Panel A*. The results in *Panel B* show the robustness of the results displayed in Figure 6a using sector-level monthly data. The results in *Panel C* reports the robustness of the results from Table 3 *Panel B* for a sample of firms that are always registered in the VAT. In all panels, the columns indicate the threshold used to winsorize the dependent variable in order to mitigate the influence of outliers. The results for p99 that are reported in the paper are presented first, then p99.9 and p95 are also shown. Standard errors clustered by sector and clustered by firms are reported for Panels A and C, and standard errors clustered by sector are reported in Panel B. For more details, see the notes of Table 3 and the notes of Figure 6. Significance levels *** 1%, ** 5%.

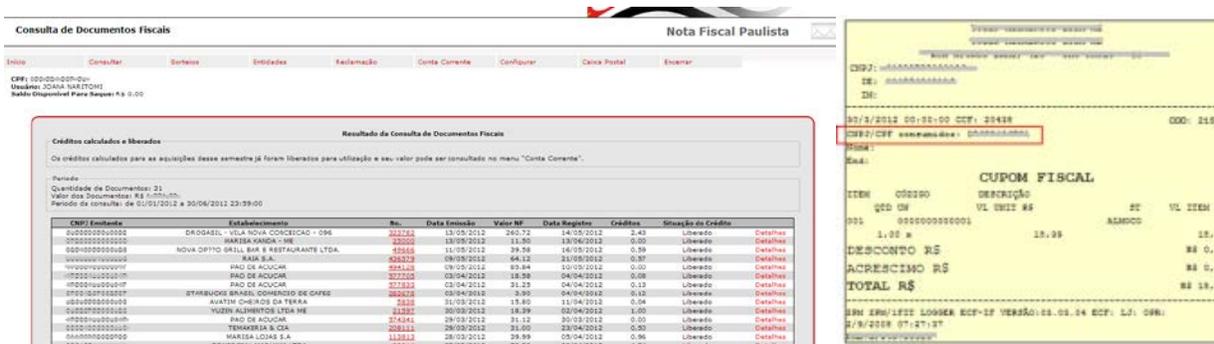


FIGURE A1: ONLINE ACCOUNT AND *NOTA FISCAL PAULISTA* RECEIPT

Note: Figures a and b are snapshots of an online account example at <https://www.nfp.fazenda.sp.gov.br>. The snapshot in Figure a is from the author’s online account. Tabs on the top of the figure can be translated (from left to right) as: *Home, Queries, Lotteries, Charities, Complaints, Current Account, Settings, Inbox, Sign out*. The tabs allow consumers to file complaints, verify whether they got a prize in a lottery, request deposits in a bank account, transfers to other enrolled consumers or transfers to charity. The interface of the account is similar to a credit card statement with a list of all receipts, the issuing date, total value of each receipt, tax rebate, and a link to the details of each receipt. Figure b shows the receipt if one clicks on the last column in Figure a for *details* of one of the purchases listed. The receipt has a field to fill in the consumers’ SSN – as highlighted in Figure b.

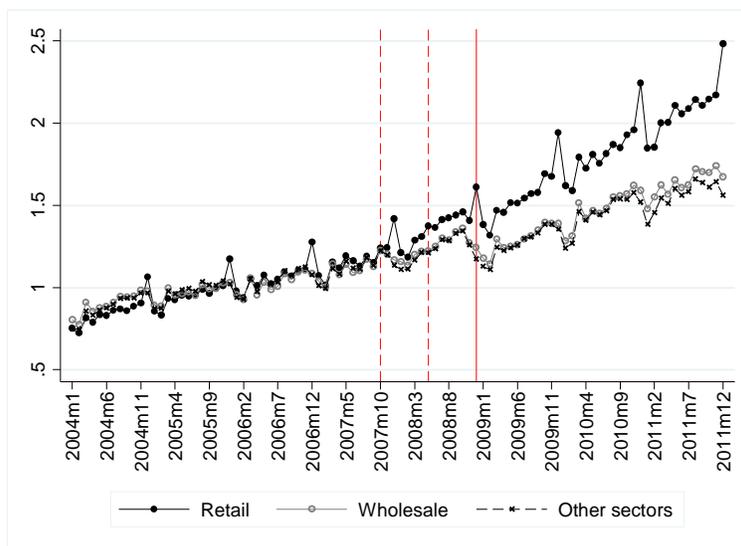


FIGURE A2: REPORTED REVENUE EFFECT – RETAIL VS. WHOLESALE VS. OTHER SECTORS

Note: Figure A2 is similar to Figure 2a: it shows reported revenue changes for retail and wholesale sectors, but it also adds all the remaining sectors as a third category. Each line is defined by the reported revenue by all firms aggregated by retail or wholesale or other sectors scaled by the average monthly reported revenue before Oct. 07 for each sector group. The figure plots the raw data, so there are spikes around December of each year follows the seasonal variation in consumption. The vertical lines highlight the key dates for the implementation of the NFP program: phase-in of sectors begins in Oct.07 and ends in May.08, and the first lottery based on the purchases with SSN receipts was introduced in Dec.2008.

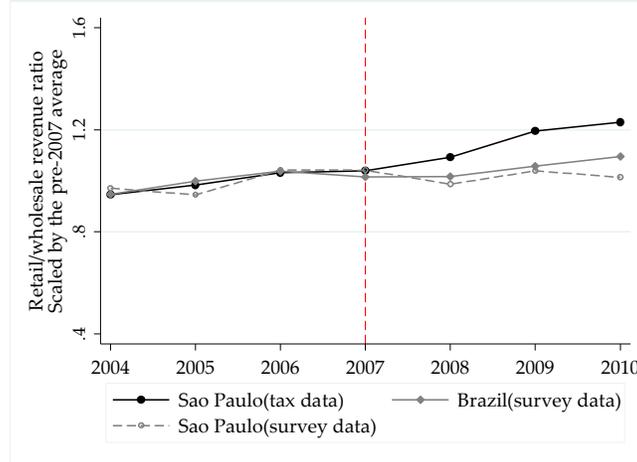


FIGURE A.3: COMPARING SAO PAULO WITH BRAZIL - CHANGES IN REVENUE RATIO

Note: The figure 3 shows changes in the retail-wholesale reported revenue ratio from the Sao Paulo tax data (black line), and changes in the retail-wholesale actual revenue ratio from a national-wide survey on the trade sector (gray lines). The dashed vertical line marks the beginning of the NFP program in 2007. The solid gray line displays the national-wide ratio (excluding Sao Paulo), and the dashed gray line shows the retail-wholesale actual revenue ratio for the state of Sao Paulo from the survey data. Each line is scaled by the pre-2007 retail-wholesale revenue ratio. The national ratio is based on the total gross revenue from sales, and retail revenue considers retail and motor-vehicles trade. Figure 3a compares changes in the revenue ratio of retail to wholesale, $r = \frac{\text{retail revenue}}{\text{wholesale revenue}}$, from the Sao Paulo administrative data to changes in the same ratio from the census survey. Each data point is scaled by the ratio r in 2004. Until the introduction of NFP in 2007, the three ratios follow similar time trends. After 2007, the ratio derived from reported revenue in Sao Paulo tax data increase, whereas the ratios derived from survey data -- in Sao Paulo state and nationwide -- remain relatively unchanged. The retail-wholesale revenue ratio was calculated from aggregate tables of the survey.

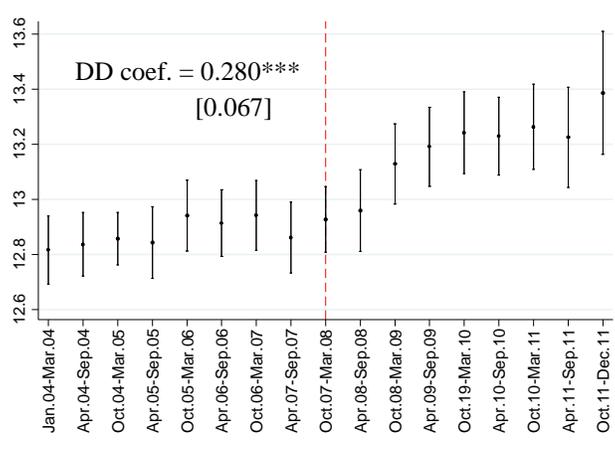


FIGURE A4: REPORTED REVENUE EFFECT – TAX SAMPLE

Note: Figure A4 is similar to Figure 2b: plots regression coefficients from estimating specification (5) using a sample of sectors for between Jan 2004 and Dec 2011. The difference in differences (DD) coefficient displayed in the figure is estimated using the specification (6) where the DD variable is defined by the interaction between a dummy for retail sectors and a dummy that equals 1 for time periods after Oct 2007. The difference with respect to Figure 2b is that it restricts attention to a sample of firms that are in sectors where there is little tax withholding and, therefore, the firm-level reported liabilities in the data best approximates their own tax liabilities (see Online Appendix B for more details).

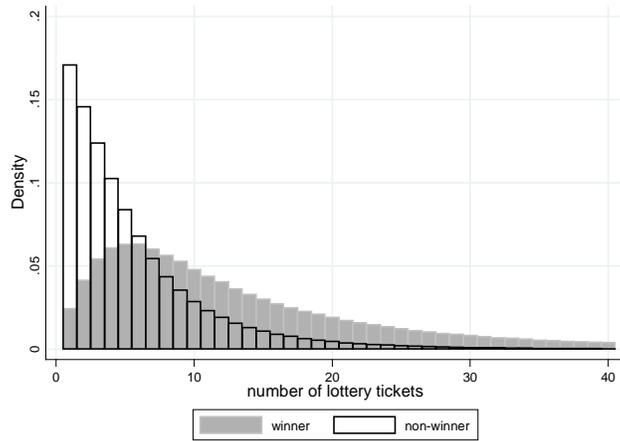


FIGURE A.5: LOTTERY TICKETS DISTRIBUTION – DECEMBER 2009

Notes: Figures A.5. shows a histogram for the number of lottery tickets winners and non-winners hold in December 2009 as an example for the re-weighting of observations for the analysis in Section 4 and described in Online Appendix B. A lottery ticket is generated for every 50 dollars a consumer spends in SSN receipts; so 50 receipts of 1 dollar or 1 receipt of 50 dollars are equivalent, and generate 1 lottery ticket. There is common support between the two groups for lottery ticket holdings below 40, and the winner group holds more lottery tickets than the non-winner group. The graphs were constructed from the consumer sample described in Section 2 and Online Appendix B.

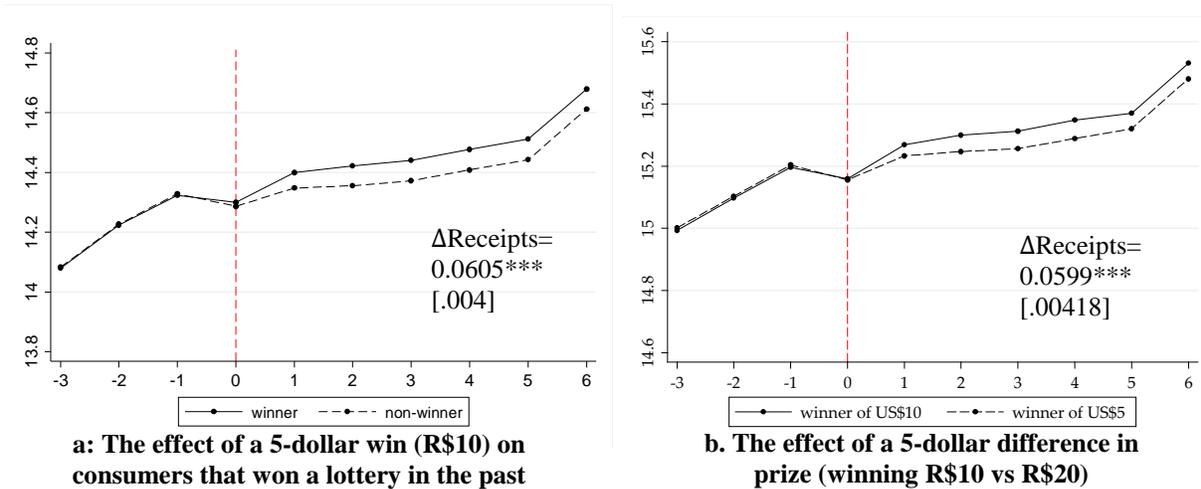


FIGURE A6: THE EFFECT OF A SMALL LOTTERY WIN – WINNERS ONLY

Note: The graphs show total raw number of receipts consumers ask by month aggregating all lotteries from June 2009 to June 2009. The x-axis is the number of months since the consumer participated in a lottery. Figure A6a shows the effect of a U.S. \$5 lottery win (R\$10) for consumers that have won a lottery once before, in which case they have already verified that the program works as advertised. Figure A6b compares winners of two different prize amounts that differ by U.S. \$5 (R\$20 vs R\$10). In this graph, both sets of consumers won a prize, so the difference is driven by the size of the prize and not a discouragement effect from not winning. Before taking the averages in each case, I create bins for each possible number of lottery ticket holdings from 1-40 tickets in each monthly lottery for 12 lotteries between June 2009 and June 2011. Then I re-weight the non-winners group such that each bin carries the same relative weight as the winner group distribution across lottery ticket holdings (see Online Appendix B for more details). The DD coefficient displayed in each graph is based on estimating specification (10) using the micro-data and the lottery ticket weights. Standard errors are clustered by lottery draw. Significance level *** 1% ** 5%.

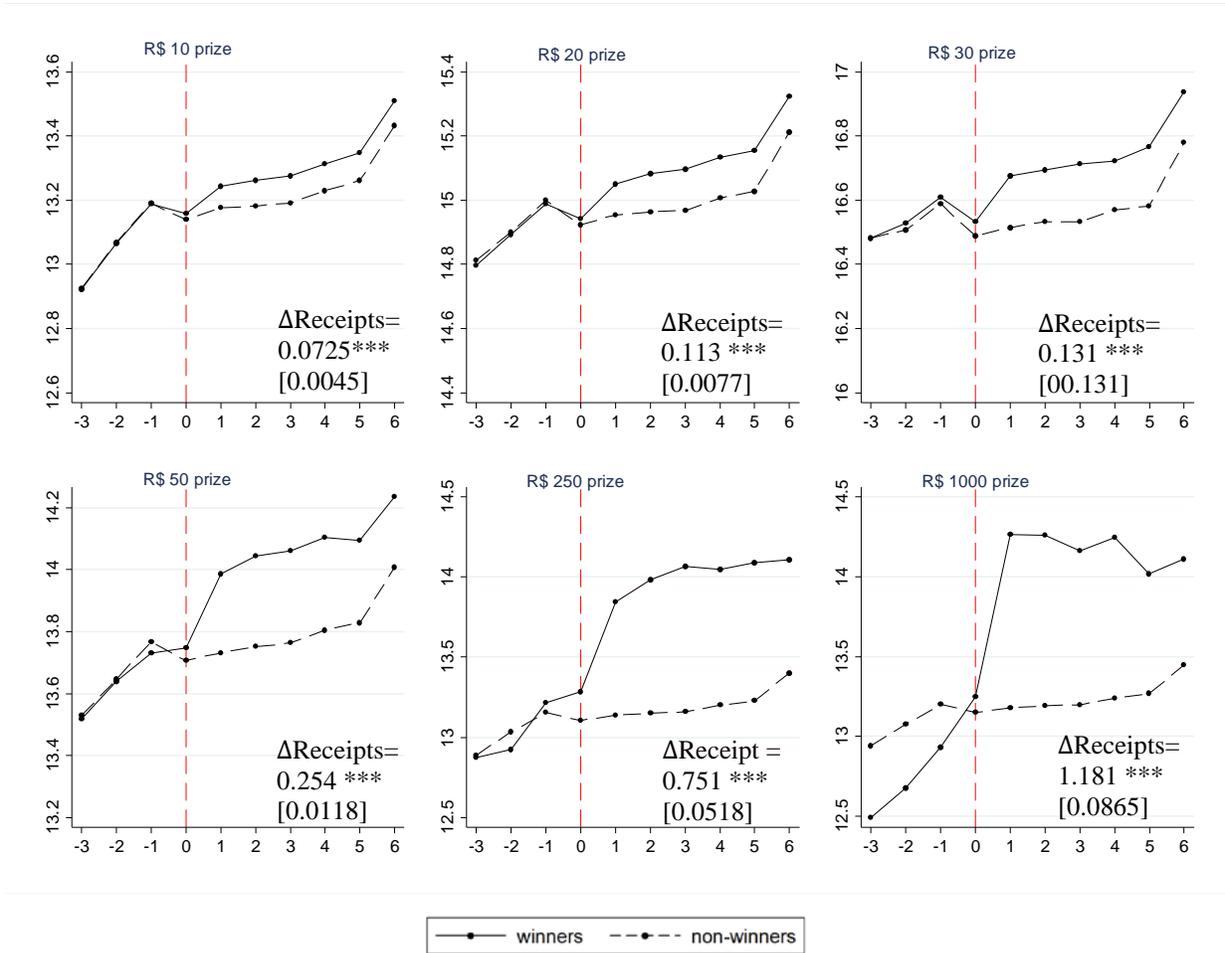


FIGURE A7: THE EFFECT OF LOTTERY WINS ON THE NUMBER OF RECEIPTS

Note: The graphs show the total raw number of receipts consumers ask by month aggregating all lotteries from June 2009 to June 2011. The x-axis is the number of months since the consumer participated in the lottery. The figures show the results for different cash prizes. In each of the lotteries there are 1,407,394 prizes of R\$10, 76,303 prizes of R\$20, 15,000 prizes of R\$50, 1,000 prizes of R\$250, and 300 prizes of R\$1000 (US\$1=R\$2). Because it is common for individuals to hold more than one lottery ticket in a month, there are many cases of consumers that get a total of R\$20 or R\$30 by winning a combination of a R\$10 and/or R\$20 prizes. Before taking the averages in each case, I create bins for each possible number of lottery ticket holdings from 1-40 tickets in each monthly lottery for 24 lotteries between June 2009 and June 2011. Then I re-weight the non-winners group such that each bin carries the same relative weight as the winner group distribution across lottery ticket holdings (for more details see Online Appendix B). The DD coefficient displayed in each graph is based on estimating specification (10) using the micro-data and the lottery ticket weights. Standard errors are clustered by lottery draw. Significance level *** 1% ** 5%.

ONLINE APPENDIX B: DATA AND RE-WEIGHTING

This appendix provides additional information on the datasets and variables discussed in Section 2, and on the re-weighting exercise from Sections 4: whistle-blower effect and lottery effect.

B.1. Firm Data

Firm panel. The data was de-identified, and a scrambled identifier was created for each establishment and firm. For the firm-level analysis in the paper, I aggregate all the data from establishments by firm. This is possible because a dataset based on the registry of firms allowed the link between the scrambled ids of establishments and firms.

The first data source are tax forms from establishments in the tax regime RPA ("Regime Periódico de Apuração") that requires establishments to file monthly a detailed tax return of all sales called GIA/ICMS ("Guia de Informação e Apuração do ICMS") to assess the total VAT due by the establishment in a given month. These tax returns include how much of sales and input transactions fall into different categories of the tax code (e.g., different rates and withholding). These are called CFOP ("Código Fiscal de Operações e Prestações). Due to confidentiality concerns, the data available to this project does not include all the transaction details by each tax code. A few aggregations of such codes were added to the data as I explain below.

The second source of data is composed by tax forms from establishments in a simplified tax regime called SIMPLES. As is common in VAT systems across the world, there is a threshold below which firms do not pay taxes over the value added. In the period of analysis, firms that have yearly gross revenue of less than U.S. \$1.2 million can choose to be in a simplified tax regime called SIMPLES in which firms pay taxes over gross revenue.

For the SIMPLES establishments I combined monthly data for establishments in Sao Paulo from three different sources: (i) tax returns from the state's *SIMPLES Paulista* in all months between 2004 and until June 2007; (ii) tax returns for the DASN-SP ("Declaração do Simples Nacional-SP") from July 2007 until the end of 2008; (iii) tax returns from DASN ("Declaração anual do Simples Nacional") between 2009 and 2011. The changes in data sources are due to the fact that there was a separate SIMPLES regime for federal and state taxes before June 2007. After that, states and federal government centralized in a single system all SIMPLES tax information, and there was a transition period in which states and the federal government kept separate records.¹

Sector. The sector definition used in the analysis is a 7-digit CNAE (Classificação Nacional de Atividades Econômicas) based on a snapshot of the registry of establishments in Sao Paulo in 2011. It was not feasible to rebuild the registry as of Oct.2007. However, the policy did not generate any incentive to change registration from a retail CNAE to a wholesale CNAE, so any change in CNAE since the policy implementation is likely to be orthogonal to the main variation used to identify the impact of the policy.

¹The datasets listed in (i) - (iii) have some months of overlap, which allowed me to cross-check the information available in each of them, and verify that these system changes did not generate mechanical changes in reporting.

Because firms are the unit of analysis in the micro-data, and sector is defined by establishment, I assigned CNAEs to firms based on the establishment CNAE. 94.08% of firms have a single establishment and among firms that are multi-establishment, most of them had multiple establishments in the same CNAE, so this process was straightforward for a vast majority of cases. 1.35% of firms have more than one establishment registered in more than one CNAE. For these firms, I assigned to the firm the CNAE of the establishment that is registered as the firm's headquarter. For 0.07% of firms, there was no headquarter indicated in the registry of Sao Paulo. In these cases, I assigned the CNAE of the establishment with most revenue during the period of analysis.

Tax liabilities measurement. As explained in Section 2 of the paper, there are measurement problems in the data for firms' tax liabilities. The variable I observe in the data is the amount of tax a firm is due to remit to the tax authority, not their tax liability. There could be substantial differences between these two quantities. An important driver of this difference is tax withholding within the VAT chain: part of the tax that is due by an firm is withheld and remitted by a upstream or downstream trade partner. If I had access to all the transaction values itemized by tax code in firms' tax returns, it could be possible, in principle, to recover the tax liability of each firm. However, this detailed data from tax returns was not available to this study for confidentiality concerns as I mentioned above.

This measurement problem introduces mechanical drops and increases in tax liabilities by firms that are difficult to control for as withholding rules are based on products and I do not observe products. Yet, there are some sectors that are less affected by withholding than others. Therefore, I restrict attention to all firms in a set of sectors with little withholding for which the reported tax due best approximates tax liabilities.

To give a concrete example: products like gasoline and some pharmaceuticals have upstream withholding policies, i.e., the producers or distributors withholds the tax for the entire supply chain. For firms that buy and sell such products downstream, the tax remitted to the government will depend on changes in withholding policies and the composition of goods that they sell or buy. Therefore, the idea is to remove from the data drugstores or gas stations as a relevant share of the trades in those sectors are affected by withholding.

In addition, backing out the tax liability from reported revenue and reported inputs is not straightforward because, depending on the tax code that applies to a given transaction, the tax liability it generates can be different. The VAT system is a credit-invoice method. With a single rate and no exemptions, it could be possible to back out the liability by simply looking at reported revenue minus reported inputs. However, this is not the case in Sao Paulo as there are multiple rates and exemptions.

To identify sectors less affected by withholding I proceed as follows: for firms in the VAT, which submit more detailed tax returns, I obtained an aggregation of the total values of input and output transactions that are in tax codes related to withholding. I aggregate these firms by sector, and calculate how much of total inputs and sales transactions are affected by withholding during the period of analysis. Then I restrict attention to sectors for which neither the input or output

transactions affected by withholding represent more than 1% of the total input or output reported by VAT firms in those sectors.

I focus on sectors instead of firms because I do not observe this withholding information for all firms, only for firms that file VAT forms. Firms in the same finely defined 7-digit sector will likely be similarly effected by withholding as they will sell a similar set of products. In addition, there are concerns about some firms making systematic mistakes in reporting withholding values in exemptions tax codes when the sale of a product does not generate a refund, in which case both tax codes would have the same null implication for the tax due.

One concern with the procedure above is that the set of sectors I look at for the tax liabilities analysis is not necessarily representative. I address this concern by allowing a direct comparison between this sample and the overall sample. In the empirical analysis (see Section 5 and Table 3) I present the results for reported revenue for the sample with little withholding alongside the results for tax liabilities to allow for a direct comparison between the this subsample of sectors and the main sample. In addition, Figure A4 in the Online Appendix A shows the main DD graph in Figure 2b for reported revenue (Difference coefficients for 6-month time bins between Jan. 2004 to Dec. 2011)

Reported inputs data. Only firms in the VAT regime (RPA regime mentioned above) file information about inputs. Therefore, in order to analyze the effect of the policy on reported inputs, I have to restrict attention to firms that are in the VAT and have not switched tax regimes during the period of analysis. The reported inputs data was extracted as a separate file. The scrambled identifiers used in this extraction were different than the ones used in the original data. However, the procedure do de-identify and clean the data was the same. The files included the GIA/ICMS forms, and the history of tax regimes of firm ids that are in these files to help identify firms that never switched out of the VAT regime between Jan 2004 and December 2011. By imposing this restriction the sample size shrinks to 44,211 firms. In the analysis I present the results for reported revenue for this sample alongside the results for reported expenses to allow for a direct comparison between the this sample and the main sample.

Receipt data. The receipt data is constructed from a dataset that has transactions with SSN-identified receipts between January 2009 and December 2011. The transaction level data is a linked establishment-consumer data. The data was de-identified, and a scrambled identifier was created for each establishment and consumer. The datasets between October 2007 and December 2008 were not available to this study. The available data restricts attention to final consumers SSN ("CPF" holders), i.e., I do not have information on receipts issued with the SSN of other establishments or charities. Also, the data on approximately 90 consumers who won one of the top 3 lottery prizes of over U.S. \$500 dollars in each monthly lottery between January 2009 and December 2011 were excluded from the dataset for confidentiality reasons. For retail firms, there is also a data with the total count of receipts issued - with or without SSN - between January 2009 and December 2011.

B.2. Re-weighting, Complaints and Lotteries

First, I introduce the re-weighting methods used in Section 4 of the paper. Then I provide further details and discussion of the sampling and empirical design of the impact of the first complaint for firms and the impact of lottery wins for consumers in Section 4.

B.2.1 Re-weighting

In both event studies I use in the paper - complaints and lotteries - I use a re-weighting method based on DiNardo *et al.* (1996) (DFL) to flexibly control for the odds of the event. The Appendix B of Yagan (2015) presents a thorough description of an application of DFL re-weighting. In this section, I explain how I use the weights in the applications of the paper, but it largely based on the description in Yagan (2015).

First, define the groups g that are being compared (e.g., consumers that are lottery winners vs. losers, or firms that received complaints vs not). Then divide all observations into bins b according to the relevant traits for the realization of the event (e.g. lottery ticket holding or probability of getting a complaint). Then use the number of observations in every group-bin to create weights so that the within-group distribution of weights across bins equals the original cross-bin distribution of weights in some base group g (e.g. lottery winners in time t_0 or firms that received a complaint in time t_0). Intuitively, DFL holds fixed the distribution of observable traits across groups by inflating or deflating the control group to match the distribution of the treatment group.

To explain the details, consider first the case of lottery wins in Section 4, there are two groups - winners and non-winners - and 10 event-time periods between $[-3, +6]$ around each event date t_0 . I DFL-reweight across 20 groups g (= 2 groups and 10 periods). I define the base group g to be the the winner group at event date t_0 . I want to compare two consumers with the same odds of winning, i.e., that hold the same number of lottery tickets. I therefore use each lottery ticket holding between 1 and 40 (to ensure common support) to bin observations into one of 40 bins b . Let b denote the bin and let g denote the group that observation j falls in. The final weight w on observation j equals:

$$w_{jbg} = \left(\frac{\sum_{j' \in b \cap j' \in g} 1}{\sum_{j' \in b \cap j' \in g} 1} \right) \left(\frac{\sum_{j' \in g} 1}{\sum_{j' \in b} 1} \right) \quad (1)$$

where j' denotes firm-year observations generally.

Using this formula, every observation j that is in the base group g will have a final weight equal to 1. Every observation that is not in the base group will received a weight smaller or greater than 1 depending on whether it is over represented or underrepresented when compared to the base group g . The first factor in parentheses ensures that within every group g , the ratio of the sum of observations for a given number of lottery tickets b to the sum of observations in any other other number of lottery ticket holding b' is identical to the corresponding ratio in for the lottery winners in time t_0 (the base group g). The second factor ensures that the sum of each groups' final weight equals the sum of that group's original weight.

For the case of complaints, I first calculate a propensity score for getting a complaint at time t_0 . Then I create the bins for firms b based on quartiles of the propensity score distribution and follow the same procedure described above to calculate the weights using formula (1). In the next two subsections I explain in more detail the construction of the data and a discussion of the role of the DFL weights in each exercise.

B.2.2 Complaints

Complaints data. 25% of firms received at least one complaint in the period of analysis. I begin defining the complaints sample by looking at firms in the retail sector that issued at least one receipt before June 2009. For each firm i I identify the time of the first complaint any of their establishments received. Then, for each complaint date e between July 2009 and June 2011, I build a panel data with 6 month window around the complaint date where firms that received a complaint at that event date are in the treatment group and firms that did not receive their first complaint by that event date are in the control group. The same firm i can be in T or C depending on the event as the control group draws from firms that did not yet receive a complaint by event e . Then, I restrict attention to firms that have positive revenue and receipt data to make sure I have a balanced panel. The combined dataset that appends all events covers the time period between Jan. 2009 and Dec. 2011, i.e., at 6 months before and after the earliest and latest first complaint respectively.

Once I create the propensity score, I construct a dataset for each month within this period, where I keep all firm that received their first complaint that month and all firms that did not receive their first complaint that month and I re-weight the no-complaints group to match the complaints group within each quartile of the propensity score distribution. I collapse each cohort of complaints by each group and "event-month" using the weights to show the raw data in Figure 3, and use the micro-data to run specification (9) using the weights.

Re-weighting. I use a propensity re-weighting method to flexibly control for the probability of getting a complaint such that I use a quasi-random component of the timing of the first complaint by matching groups that have similar odds of getting a complaint. I estimate a propensity-score of a firm receiving the first complaint for every month-year between July 2009 and June 2011 based on pre-event characteristics. Then I use quartiles of the propensity score to re-weight firms that did not received their first complaint in the given month-year to compare with firms that received their first complaint in that month-year.

I perform this re-weighting exercise separately for each period between June July and June 2011. For each case, I restrict attention to the sectors that had at least one firm that received a complaint in a given date and I draw a 10 percent random sample of firms that did not receive the first complaint on that date to build the no-complaint sample. This sample includes both firms that did not receive their first complaint in a given date and firms that did not receive any complaint by Dec. 2011. The propensity score is estimated using a logit model on time specific trends for each sector, age of the firm, number of establishments by firm, dummies for legal nature of the firm, sector fixed effects, dummy for location in the metropolitan region of Sao Paulo, and the three lags

of third-order polynomials of reported revenue, total number of receipts issued, total number of SSN receipts issued and total number of consumers.

B.2.2 Lottery wins

Lottery data. The lottery sample covers consumers that hold fewer than 40 lottery tickets in a given month for 12 lotteries between July 2009 and June 2011. In order to perform the empirical exercises on the effects of lottery wins on consumer participation in Section 4.2 I merge this data with the *receipts data* described above. The combined dataset of lotteries and receipts covers the time period between January 2009 and December 2011, i.e., 6 months before and after the first and last lottery. I balance the panel of consumer participation and replace missing values by zero for the two key variables I use *number of receipts* and *total value of receipts*.

Consumer sample. This sample only includes consumers that participated in at least one lottery between July 2009 and June 2011. For each monthly lottery, I restrict attention to lottery winners and 10% random sample of consumers who did not win the lottery in each draw. Table 2 creates summary statistics for a balanced panel of individual SSNs with a count of the number of receipts they ask, the total value of receipts, lottery tickets and lottery prizes. The data includes zeros in months where no receipts are asked or the individual does not participate in lotteries.

For the lottery analysis, I take the winners and the sample of non-winners of each lottery, and I append all lotteries. I create a balanced sample for three months before and six months after the lottery draw, and the lottery draw month. The same consumer can be a winner a non-winner depending on the lottery. The estimation use individual-lottery FE in order to only use variation within individual-lottery, and also include calendar time FE and event-time FE.

Re-weighting. Since the number of lottery tickets is determined by the total value of a consumer's purchase 4 months before the lottery draw, the more a consumer participates in the program by asking for receipts, the higher are the odds she will get a prize in a given lottery. Therefore, it is important to carefully control for the odds of winning a prize in order to study the effect of lottery wins on consumer participation. As I describe in Section 4.2, I use the DFL re-weighting method to flexibly control for the number of lottery tickets individuals hold to ensure I use the random component of the lottery by matching the two groups based on the odds of winning prize.

Figure A.5 shows an example of the distribution of lottery ticket holdings among winners and non-winners in monthly lotteries. I create bins for each possible number of lottery ticket holdings up to 40 tickets, which is the set of lottery tickets for which there is common support between the two groups for every lottery holding. In the case of prizes that are only possible by winning a combination prizes – e.g., a U.S. \$15 total prize is always a result of winning a US\$5 prize and a US\$10 prize -, I restrict attention to lottery ticket holdings between 2 and 40 tickets. I then re-weight the non-winners group such that each bin carries the same relative weight as the analogous bin in the winner group distribution across lottery ticket holdings.

I perform this re-weighting exercise separately for each lottery win I study in Section 4.2. I construct a dataset for each prize level as described in *lottery data* above, where I keep all consumers that won a given prize (winners) and all consumers that do not win any prize (non-winners). When I compare the effect of different prize values, the pool of non-winners is the same in each lottery across the datasets I create for each prize level but they are re-weighted differently depending on the prize I am considering since the winners group of different prize levels may have slightly different distributions of lottery ticket holdings.

Once I calculate these weights for each monthly lottery, I append the panels of consumers and 10 event-time periods between $[-3, +6]$ around each event date t_0 . Then, I collapse the data by event-time using the weights for the graphs in Figure 5, Figure A5 and Figure A6. re-weight the non-winners group such that each bin carries the same relative weight as the analogous bin in the winner group distribution across lottery ticket holdings.

ONLINE APPENDIX C: CONCEPTUAL FRAMEWORK - EXTENSIONS

In this section, the goal is to extend the discussion of the conceptual framework to further discuss relevant policy dimensions: (i) imperfect take up and (ii) collusion costs.

C.1. Imperfect take up

Consider the same conceptual framework as in Section 1, where the government provides a targeted incentive α for consumers to ask for receipts. As before, suppose consumers derive benefit $\kappa(\alpha)\tau y$ from the governments' incentives to ask for receipts, but allow consumers to be heterogeneous in their participation cost ϕ_i . Let $\phi_i = \underline{\phi}$ for a share $\lambda \geq 0$ of the N consumers, and $\phi_i = \bar{\phi}$ for a share $(1 - \lambda)$, where $\underline{\phi} < \bar{\phi}$.

If $\kappa(\alpha)\tau y \geq \bar{\phi}$, all consumers will take up the policy, i.e., will ask for receipts, and the problem will be the same as the compliance decision with consumer monitoring in Section 1. In order to highlight the relevance of take-up, let $\kappa(\alpha)$ be such that $\underline{\phi} < \kappa(\alpha)\tau y < \bar{\phi}$. Therefore, a share $\lambda > 0$ of consumers will respond to the incentives and ask for receipts. Imperfect take up will weaken the enforcement effect of the policy. I will consider two alternative responses of firms: selective reporting and collusion.

Selective Reporting. Suppose firms can strategically choose how each of the N transactions will be reported to the government. In this case, the degree to which the policy can affect firms at all may depend on the baseline compliance. Let λ^* be such that $\lambda^*\bar{Y} = Y^*$. If $\lambda \leq \lambda^*$, the share of transactions for which consumers ask for receipt will be smaller than what would be reported in the absence of the policy $\lambda\bar{Y} \leq Y^*$. In this case, the receipts consumers begin to ask would be entirely infra-marginal if firms can selectively under-report transactions for which no receipt was asked, and report transactions for which consumers did ask for receipts as there is a higher detection probability if firms misreport these transactions.²

Thus, if consumer take up is sufficiently low, compliance may not change irrespective of the size of rewards. Moreover, the availability of whistle-blower channels absent of rewards would make no difference as consumers would not observe evidence of evasion. This stylized case emphasizes how important take up can be if firms do not under-report the entirety of their sales and can do selective reporting.³ In addition, the rewards offered by the government would be infra-marginal as no additional sale would be reported as a result of the policy. Thus, the baseline level of compliance matters for the cost of the policy: in the limit, if the baseline compliance close to perfect, rewards could be entirely infra-marginal even with large take-up.⁴

²This is similar to the argument used in Kleven *et al.* (2011) to describe reporting incentives for incomes that differ in third-party reporting coverage: because the tax rate and penalty are the same across different revenue items, this strategic selective reporting would be the optimal sequence for the taxpayer. The detection probability would have a S shape: it sharply increases once firms start underreporting income subject to third-party reporting.

³If selective reporting imposes a concealment cost, the policy could have an effect on compliance by increasing the cost of evasion.

⁴Similarly, if there is a share of consumers that require receipts even absent of the a reward policy (if they have a negative ϕ_i because they value compliance directly), their effort to ensure firms issue receipts would only affect firms if the number of consumers with negative costs is high enough.

If $\lambda > \lambda^*$, the consumer policy will affect the reporting decision of the firm. If firms continue reporting any transaction for which consumers ask for receipts, the reported revenue would be given by $Y = \lambda \bar{Y}$. In this case, the higher the take up, the larger the amount of sales reported by firms, so firms would mechanically increase the amount of reported sales depending on the number of consumers that take up the policy.

Since take-up depends on $\kappa(\alpha)\tau y$ being higher than the participation cost, goods with high τ and high price y would be relatively more effected by the policy as it is more likely that the reward will be higher than the participation cost for a larger share of the population. It also highlights that if participation cost are very high (e.g., $\kappa(\alpha)\tau y < \phi$), the policy may have no effect even if α is generous. Note that under selective reporting, the whistle-blower channel would have no effect as firms would not reveal to consumers evasion information. In terms of government revenue, the policy would be net positive if $\frac{\lambda - \lambda^*}{\lambda^*} > \frac{\alpha}{1 - \alpha}$.

Collusion policy. Suppose firms could offer the collusion deal to consumers that request receipts: a discount of $\kappa(\alpha)\tau y$. As before, assume that consumers take any offer that matches their valuation $\kappa(\alpha)$ of the policy. Because only consumers that ask for receipts learn about the evasion decision of firms, the risk of whistle-blowers will depend on take up as well. The detection probability would be affected less than in the case of full take up: $d_c^\lambda = 1 - (1 - d)(1 - \varepsilon)^{\lambda N} < d_c$ as there are fewer potential whistle-blowers, so $p_c^\lambda = a(E)[1 - (1 - d)(1 - \varepsilon)^{\lambda N}]$. Firms will have to transfer $\lambda(\bar{Y} - Y)\kappa(\alpha)\tau$ to consumers. Firms choose Y to maximize:⁵

$$(\bar{Y} - \tau Y)(1 - p_c^\lambda) + [(1 - \tau)\bar{Y} - \theta\tau(\bar{Y} - Y)]p_c^\lambda - \kappa(\alpha)\tau\lambda(\bar{Y} - Y) \quad (2)$$

As mentioned above, under the new policy, firms have to transfer part of the evasion rents to consumers through discounts. An interior optimal solution Y_λ^{**} satisfies the first order condition $d\pi/dY = 0$:

$$[a + a'(E).E]d_c^\lambda(1 + \theta) = 1 - \lambda\kappa(\alpha) \quad (3)$$

For the comparative statics, c can be re-written as $c = \frac{1 - \lambda\kappa(\alpha)}{d_c^\lambda(1 + \theta)}$. As before, changes in c translate into comparative statics for evasion E . The main difference under imperfect take-up is that a lower λ decreases the change in detection ability and the transfers to consumers, so $Y_\lambda^{**} < Y^{**}$. Compliance, thus, increases with the share of consumers λ that take up the policy.

Similarly to the case of selective reporting, take-up depends on $\kappa(\alpha)\tau y$ being higher than the participation cost, so goods with high τ and high price y could be relatively more effected by the policy. But in the collusion case, any share $\lambda > 0$ would change the evasion decision of the firm as firms are underreporting all transactions by the same amount, instead of selectively reporting some transactions in full and evading the remaining ones. In terms of government revenue, the net effect will be positive if $\frac{Y_\lambda^{**} - Y^*}{Y^*} > \frac{\alpha}{(1 - \alpha)}$.

⁵I assume that if the firm is audited the government will consider as tax evasion the amount not reported based on the posted price \bar{y} , not the discounted price. Therefore, \bar{Y} will be the true revenue of the firm, instead of the revenue net of transfers to consumers.

The data available to this study does not allow to distinguish between these two strategies of misreporting: selective underreporting and collusion. However, these two cases are helpful to emphasize the relevance of take up and participation costs in affecting the effectiveness of the policy.

C.2. Collusion costs

Another mechanism that is consistent with the evidence is that there is a fixed cost in setting a collusive deal with consumers that acts as a concealment cost, so the larger the number of consumers a firm interacts with, the more this policy may increase compliance as it dis-proportionally affects firms that need to collude with a large number of consumers.

In the paper, I assume that when firms make an offer to a consumer, they pay a fixed cost $\rho > 0$. In order to focus on the effect of collusion costs, assume perfect take up and homogeneous consumers. The fixed cost can be thought of as a concealment cost paid to collude at each transaction. If the firm follows a collusion policy with all its consumers, the total cost of collusion ρN would affect the extensive margin decision between evasion and full compliance, but not the intensive margin of compliance as discussed in the paper.

Given the fixed cost ρ , if firms can optimize the number of collusive deals and offer discounts to some consumer but not others, they could restrict collusion to a smaller set of consumers. The total number of collusive deals a firm makes is given by the amount evaded $E = \bar{Y} - Y$ divided by the value of an individual transaction \bar{y} . Thus, the total negotiation cost is increasing in the number of collusive deals a firm makes given by $(\frac{\bar{Y}-Y}{\bar{y}})$. Because firms will only reveal evasion information to a subset of consumers, $\hat{d}_c = 1 - (1-d)(1-\varepsilon)^{\frac{\bar{Y}-Y}{\bar{y}}} < d_c$ as there are fewer potential whistle-blowers. So $\hat{p}_c = a(E)[1 - (1-d)(1-\varepsilon)^{\frac{\bar{Y}-Y}{\bar{y}}}]$

$$\pi = (\bar{Y} - \tau Y)(1 - \hat{p}_c) + [\bar{Y}(1 - \tau) - (\bar{Y} - Y)\theta\tau]\hat{p}_c - (\bar{Y} - Y)\kappa(\alpha)\tau - (\frac{\bar{Y} - Y}{\bar{y}})\rho \quad (4)$$

$$[\hat{p}_c - \hat{p}'_c(E) \cdot E](1 + \theta) = 1 - \kappa(\alpha) - \frac{\rho}{\tau\bar{y}} \quad (5)$$

The marginal benefit of evading an extra dollar is reduced further by $\frac{\rho}{\tau\bar{y}}$. The total change in the marginal benefit of evasion amount is equal to the transfer firms need to make to consumers in the collusive deal plus the cost per dollar negotiated with consumers. Therefore, the costs of collusion enter as an extra penalty for each dollar evaded.

Here, it becomes more explicit that the value of the transaction \bar{y} will matter for this policy as firms with the same level of evasion may be affected differently as firms with a low ticket items would have to collude with a higher number of consumers for a given total revenue \bar{Y} . Since the number of potential whistle-blowers depend on how many collusive deals are done, \hat{p}_c is also higher for lower \bar{y} , further pushing firms towards compliance.

Therefore, the size of the transaction could matter through collusion costs and the higher risk of whistle-blowers for a given firm size. It is worth noticing that predictions about the size of the transaction should also depend on take up. If take up is imperfect as the cost of asking for receipts

is too high for some consumers, fewer consumers would be willing to ask for receipts for small ticket items. In the data, the effect is stronger for low ticket items, which is consistent with the transaction cost for consumers not being very high and the collusion costs (through fixed costs and/or additional whistle-blower threat from volume of transactions).

ONLINE APPENDIX D: EFFECTS ON FORMAL EMPLOYMENT AND NUMBER OF FIRMS

The observed enforcement affect generated an increase in the effective tax rate. This change may affect firms that were on the margin of exiting the market, entering the market or firing employees. In this appendix, I analyze the implications of the policy on formal employment and the number of firms.

D.1. Formal employment

Employer-employee data. From the Brazilian Department of Labor, I use annual reported employment for all formal establishments in Brazil (RAIS/CAGED). The data covers all formal establishments that have at least one formal employee. All formal firms must report to the Department of Labor their employment information in a yearly basis. I use a version of this data that aggregates the total number of employees by 7-digits sector definition. Because the data from the Department of Finance of Sao Paulo is de-identified, it cannot be matched with the employer-employee dataset (RAIS/CAGED). In order to analyze the impact of the program on employment, I construct a panel of CNAE-year for each Brazilian state.

Employment effect. I employ two strategies. First, I apply the same difference in differences (DD) design in equation (6) of the paper comparing retail and wholesale sectors. Because the employment data is measured yearly I define *after* as $t \geq 2007$. Figure D1 shows the DD coefficient from estimating a specification similar to (6) in the sector yearly panel with standard errors clustered by sector. It also shows the point estimate and 95% confidence intervals for the coefficients of a more flexible DD specification in (5) using yearly dummies interacted with the retail dummy, and using the 2007 interaction as the omitted category.

Then, because this sample covers the entire country, I can use retail-wholesale difference in Sao Paulo and compare with the retail-wholesale difference in other states in a triple difference in differences (DDD).

$$\ln Empl_{stm} = \pi_m + \eta_s + \lambda_t + \psi SP_m \cdot Post_t + \alpha Treat_s \cdot Post_t + \gamma SP_m \cdot Treat_s + \beta Treat_s \cdot Post_t \cdot SP_m + u_{stm}$$

Where $\ln Empl_{stm}$ is the natural log of total formal employment in sector s , year t and state m . π_m is state fixed effects, η_s is sector fixed effects, λ_t is year fixed effects. Figure D1.B. shows the coefficient β estimated using the above specification, and also displays the point estimate and 95% confidence intervals for a more flexible specification where the interactions time is done with year dummies separately for each year, using 2007 as the omitted category.

Even though the point estimate is negative in both cases, the coefficients are not statistically different than zero, suggesting that the policy, on average, had no effect on formal employment. The fact that I find no or slightly negative effect on employment is consistent with the increase in

reported revenue being a reporting effect, rather than an actual increase in sales, in which case I could potentially observe an increase in employment.

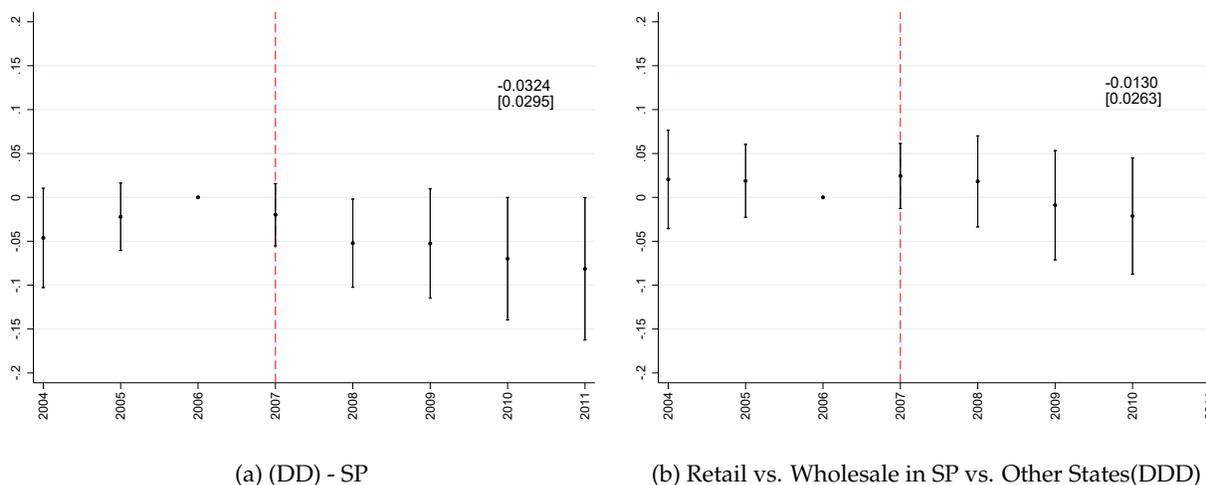


FIGURE D.1: LOG OF NUMBER OF FORMAL EMPLOYEES

Note: The figure plots the log of number of firms in retail vs. wholesale. The count of number of firms considers all firms that reported positive revenue each year. The figure displays the DD coefficient from estimating a specification similar to equation (6) in a 7-digit sector yearly panel with 1,680 obs., and using the log of total number of firms as the outcome. Standard errors are clustered at the 7-digit sector level.

D.2 Number of firms

The increase in enforcement could affect both entry and exit decisions. To assess whether these margins were affected by the policy, I use the log of total number of firms observed in each year and sector as a simple measure of potential changes in entry and exit decisions.

Number of firms To calculate the number of firms, I define as the year of exit the last year a firm had any sales during the period of analysis. Firms may submit forms with zero activity due to the slow process of closing a firm in Brazil, so defining exit as when they stop reporting positive economic activity would better capture the timing of exit. The results are robust to alternative ways of counting firms, where I count any firm that files a tax form at least once each year even if it has zero sales or inputs.

Figure D.2 plots the log of number of firms in retail and wholesale sectors. The figure also displays the DD coefficient from estimating a specification similar to equation (6) in the paper but using the log of number of firms as the outcome. The coefficient is not statistically distinguishable from zero, which indicates that, on average, the policy did not affect the number of firms in retail sectors differently than in wholesale sectors.

The evidence above indicates that the increase in tax enforcement did not affect employment or the number of firms of affected sectors. The null effect indicates that the implied increase in the effective tax rate may not be large enough to affect firms along these margins, and may just

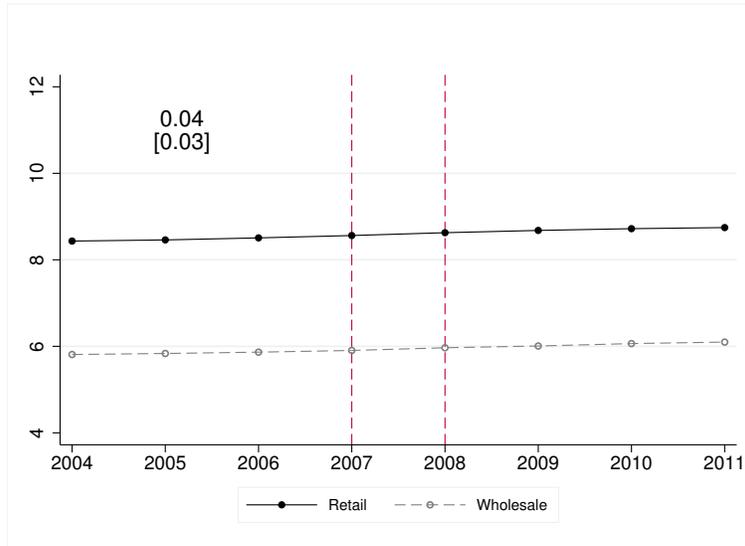


FIGURE D.2: LOG OF NUMBER OF FIRMS

Note: The figure plots the log of number of firms in retail vs. wholesale. The count of number of firms considers all firms that reported positive revenue each year. The figure displays the DD coefficient from estimating a specification similar to equation (6) in a 7-digit sector yearly panel with 1,680 obs., and using the log of total number of firms as the outcome. Standard errors are clustered at the 7-digit sector level.

reduce evasion rents.⁶ However, it is possible that changes may occur after the period of analysis. Another possibility is that firms can potentially adjust other margins that I do not observe in the data such as, for instance, informally-hired workers.

⁶Moreover, establishments may be able to pass-through this tax increase to consumers. Data on prices and quantities – which have not been available for this project – would be needed to understand the incidence of the policy.

References

- DINARDO, JOHN, FORTIN, NICOLE M, & LEMIEUX, THOMAS. 1996. Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach. *Econometrica*, **64**(5), 1001–1044.
- KLEVEN, HENRIK JACOBSEN, 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.
- YAGAN, DANNY. 2015. Capital tax reform and the real economy: The effects of the 2003 dividend tax cut. *American Economic Review*, **105**(12), 3531–63.