Loss Aversion, Transaction Costs, or Audit Trigger? Learning about Corporate Tax Compliance from a Policy Experiment with Withholding Regime

Carrillo, Paul and Emran, M. Shahe

16 June 2018

Loss Aversion, Transaction Costs, or Audit Trigger? Learning about Corporate Tax Compliance from a Policy Experiment with Withholding Regime[1]

Paul E. Carrillo
George Washington University

M. Shahe Emran
IPD - Columbia University

This Version: June 2018, Comments Welcome

ABSTRACT

We analyze firm's tax choices facing a withholding and enforcement regime with a focus on three mechanisms of bunching: (i) transaction costs, (ii) withholding threshold as a reference point for taxpayer that creates a kink due to loss aversion, and (iii) withholding threshold as a reference point for audit (audit trigger model). The transaction costs model predicts that none of the firms that bunch at the withholding threshold would declare higher taxes when withholding rate is increased, as was the case in Ecuador in 2007. Evidence from a triple-difference research design shows higher tax payments by these firms. A prospect theoretic model with the power value function of Kahneman and Tversky (1979) does not generate bunching at the withholding threshold. While linear prospect theory (LPT) can generate bunching under certain conditions, it also yields testable predictions that are not consistent with the behavior of a significant proportion of firms. Under the LPT, given an enforcement and withholding regime, if a firm bunches in one year it should also bunch in all the following years, or if it unbunches in a following year, it should declare taxes *less* than the withheld amount. The evidence from panel data on the universe of all corporations in Ecuador shows very low persistence in bunching: conditional on bunching at least once, only 3-4 percent firms bunch every year before changes in the withholding rate, and among the firms that unbunch 35-40 percent declare taxes *more* than the withheld amount, thus contradicting the LPT model for a substantial proportion of the firms. Using the Sasabuchi t test as developed by Lind and Mehlum (2010), we find that the relation between probability of bunching and assets of a firm is inverted-U which is consistent with the audit trigger model. The evidence suggests that the behavior of the firms cannot be captured by a single model. The strength of enforcement is important in determining bunching in an LPT model which suggests cross-country differences in the role played by loss aversion in bunching of taxpayers at policy thresholds.

Keywords: Loss Aversion, Reference Dependence, Transaction Costs, Audit Trigger, Bunching, Withholding, Firms, Profit Tax, Tax Evasion, Ecuador

JEL Codes: H25, H26, O23, O12

---

1. Introduction

Income tax withholding at the source by an employer has a long history, and it is now a prominent feature of tax administration in almost all developed and developing countries.[2] Recent work by Saez et al. (2016) has underscored the importance of withholding at source and third-party reporting in understanding the puzzle of "over compliance" in modern tax administration. While from its inception, withholding as a tax enforcement tool has primarily been applied to employment income, in many developing countries withholding has become an important instrument for ensuring compliance by the small firms and self employed where a fixed percentage of a firm's sales are withheld at the time a transaction is made, and then transmitted to the tax authority.[3] Unlike other widely-discussed features of tax system such as tax brackets and VAT exemption, withholding does not affect tax liability, and thus is irrelevant for the decisions of taxpayers in the standard Allingham-Sandmo (1972) model. However, withholding may affect a firm's tax compliance because it creates a threshold. The observation that thresholds in the tax system can affect tax reporting and distort economic decisions by creating kinks and notches has been the focus of a growing theoretical and empirical literature (for excellent surveys, see Slemrod (2016), Kleven (2016)).

We make two contributions to this literature. First, we develop models and derive testable implications of a firm's tax choices faced with a withholding and enforcement regime that focus on three mechanisms of bunching: (1) transaction costs, (2) withholding threshold as a reference point for audit, and (3) withholding threshold as a reference point for the taxpayer that generates a kink in the preference due to loss aversion. Second, we use rich panel data on the universe of corporations from Ecuador with quasi-experimental variation in withholding rates to discriminate among alternative models. The focus of the literature has been on what we can learn from the bunching of firms at the withholding threshold (more broadly, at a policy threshold); we show that much can be learned from the unbunching behavior of the firms that

---

[2]Withholding or "taxation at source" was introduced first in England when income tax was reenacted in 1803. The principle of "taxation at source" was designed to tackle "the gross fraud and evasion" that characterized the first income tax introduced in 1799 (see Sabine, 1966).

[3]Tax withholding and various forms of presumptive taxation have been adopted in many developing countries because of the difficulties in tax enforcement with large informal sector and weak information base (Emran and Stiglitz (2005, 2007), Soos (2003)).

bunch in one year but not in another, without any change in the enforcement regime or the withholding rate (more broadly, without changes in the rules of the game).

Since the withheld amount can act as a natural reference point when a firm decides its tax declarations, we develop a model of tax evasion with reference dependence, loss aversion, and diminishing sensitivity. Withholding cannot generate bunching of tax payments in a Prospect theoretic model with standard S-shaped power utility function of Kahneman and Tversky (1979). Bunching due to loss aversion and reference dependence can occur in a linear prospect theory (henceforth LPT) model where preference is piece-wise linear in both the losses and the gains domains under the following conditions: (i) there exists a subset of firms with true tax liability higher than the withheld amount, (ii) the strength of tax enforcement falls in an intermediate regime, (ii) the degree of loss aversion is high enough, given an enforcement regime. An important implication is that the likelihood of bunching by taxpayers due to loss aversion and reference dependence will vary across countries because of differences in the tax administration, and the probability of bunching due to tax evasion is likely to be lower in developing countries with weak tax administration. We develop a set of testable predictions implied by kinked linear preference that focuses on the intertemporal behavior of the economic agents who bunch at the withholding threshold, and take advantage of the panel data to test them.

The possibility that the withholding threshold may be a reference point primarily for tax enforcement, instead of the taxpayer, has received relatively little attention in the literature. A withholding threshold may affect the monitoring of firms by tax authority in a discontinuous fashion, because tax devolutions (when tax liability is lower than withheld funds) usually result in higher audit probabilities for a firm in many developing countries.[4] We extend the Allingham-Sandmo model to incorporate differential audit probability across the withholding threshold (called the Audit Trigger model and denoted as the AT model for short), and show that the audit trigger may induce firms to bunch the tax payments at around the withholding threshold.[5]

A third mechanism behind bunching is transaction costs: some firms that expect positive

---

[4]As noted by Andreoni et al. (1998), many tax agencies follow a cut-off rule, and concentrate their audit resources below the cut-off. The withholding threshold may act as a focal point for establishing such a cut-off for auditing the firms.

[5]While the possibility that a refund claim can increase audit probability has been noted in the literature, to our knowledge, there is no formal model.

refunds may forgo the refund if the transactions costs of claiming refund is large enough. In contrast to prospect theory and audit trigger models, bunching in this case does not indicate tax evasion by firms.

For our empirical analysis, a central identification challenge is that, in cross-section surveys, it is difficult to identify bunching that is due to evasion rather than over-reporting, and to isolate them from lumps in the firm distribution caused by technological and economic factors unrelated to withholding. Withholding requirements in Ecuador vary across industries and over time: for some industries (manufacturing, construction and wholesale trade and retail) they increased from 1 to 2 percent in June 2007; for other industries (transportation) they remained unchanged at 1 percent. The evidence presented below shows that bunching around the withholding rate (of 1 percent) is clearly observed in 2004-2006, i.e., before the policy change. More important is the finding that the pattern radically changes in 2007 only for those industries where withholding rates were raised, implying that bunching is primarily due to optimizing choices of the firms.[6]

A testable implication of the null hypothesis of bunching due to transaction costs is that the firms situated around the 1 percent threshold before the policy reform should not increase their tax payments in 2007. The evidence from a triple-difference research design is opposite: these firms *increased* their tax payments in 2007. The higher tax payments observed in 2007 for the firms located at around the 1 percent threshold in 2006 is, however, consistent with both the audit trigger (AT) and linear prospect theory (LPT) models.

The LPT (Linear Prospect Theory) model of evasion implies that bunching by firms at the withholding threshold is extremely persistent over time if the enforcement regime and the withholding rate remain unchanged. In this model, an initially-bunched firm can unbunch in the following years only if it experiences a specific type of shock switching its true refund from negative to positive.[7] Our analysis shows that such a firm will declare tax payments to the *left*

---

[6]The withholding rate was changed in July 2007. For the affected firms, during the first half of the year 2007 the withholding rate was 1 percent, while during the second half it was 2 percent. Thus one should not expect any bunching around 1 percent or 2 percent thresholds for the year 2007. Also, if bunching around a new threshold requires adjustments, for example, in accounting data, then there can be slow response, and one should observe a similar bunching only after some years of the change in the threshold. Before 2007, the withholding rates remained unchanged since 1989.

[7]In the LPT model, a firm's tax payments can bunch at the withholding threshold only if its true tax refund is negative, i.e., the true liability is higher than the withheld amount.

of the withholding threshold. In contrast, under the null hypothesis that the observed bunching in a given year is due to the change in the audit rate across the threshold, an initially-bunched firm can unbunch irrespective of whether the shock to profits is positive or negative, and some of the unbunching firms are likely to declare taxes higher than the withheld amount. We use a rich panel data set for 2004-2006 when the withholding rate and the enforcement regime were unchanged to test the null hypothesis of persistence due to loss aversion in an LPT model. The evidence shows that, among the firms that bunch at least once, only 3-4 percent of firms bunched in all three years (2004, 2005, 2006) before the increase in the withholding rate in 2007. This is inconsistent with the prediction of strong intertemporal persistence in bunching. Among the firms that unbunch, 35-40 percent declare taxes more than the withheld amount which is impossible under the LPT model, but is consistent with the AT (Audit Trigger) model. The evidence thus suggests that the behavior of a subset of firms is not consistent with the widely-used linear preference prospect theory model with loss aversion and reference dependence.[8]

We provide additional evidence consistent with the AT (Audit Trigger) model by exploiting a substantial body of theoretical literature and empirical evidence that firm size (measured in terms of assets) is negatively correlated with risk averseness of a firm (Herranz et al. (2015), Greenwald and Stiglitz (2013)). Under the null hypothesis that the audit trigger model is correct, the firms in the middle of the asset size distribution are the most likely to bunch, implying that the relation between the assets and propensity to bunch is inverted U shaped. In contrast, only the largest firms bunch under the prospect theory model, because linear preference is likely to be a good approximation for the largest firms with diversified portfolio. We find that the probability of bunching is marginally *smaller* for the large firms when compared to the other firms, and the estimate is statistically significant at the 1 percent level. A formal test using a method developed by Lind and Mehlum (2010) that can discriminate between a monotonic concave relation and an inverted-U relation finds evidence consistent with the audit trigger model.

---

[8]Our finding that, for a substantial proportion of firms, bunching at the withholding threshold in Ecuador is not likely to be due to loss aversion is different from the recent evidence in the context of Sweden and USA that finds an important role for loss aversion for *individual* taxpayers (Engstrom et al. (2015), Rees-Jones (2017)). Our theoretical results offer a plausible explanation for this difference: the relatively weak tax administration in Ecuador makes it difficult to satisfy the conditions required for bunching caused by loss aversion and reference dependence noted above.

The analysis presented below has important implications for the broader literature on bunching of economic choices at policy thresholds. First, our analysis adds a caveat to the interpretation of the parameter estimated from observed bunching as an estimate of degree of loss aversion; such an interpretation requires credible evidence in favor of linear preferences on both sides of the threshold. The implications of kinked linear preference developed in this paper can be useful in a variety of other contexts, and can be tested using panel data. Second, for a proper understanding of the observed choices, it is necessary to develop testable predictions from alternative models based on reference dependence on either side of an economic relation (tax payers vs. tax administration in our case). While the behavior of many firms is consistent with loss aversion and reference dependence in taxpayer's optimization, there is a substantial proportion of firms that behaves more according to the audit trigger model, where the withholding threshold is a reference point for audit, but the taxpayers follow the Allingham-Sandmo model. Thus relying on a single model may not be appropriate for understanding the aggregate behavior. Third, whether loss aversion leads to bunching in economic choices depends on the strength of enforcement, implying heterogeneity across countries and contexts.

The rest of the paper is organized as follows. Section 2 discusses the related literature to situate the concerns and contributions of the paper in context. Section 3 develops alternative models of firm's tax evasion with a focus on the implications of withholding. The next section (section 4) is devoted to empirical strategy and lays out the testable implications of the theory for bunching and unbunching of the firm's tax payments at around the withholding threshold. We discuss the institutional details about Tax Administration in Ecuador in section 5. In section 6, we discuss the data and then present the empirical results showing that there is excess bunching at the withholding threshold and it is not driven by lumps in firm characteristics. Section 7, arranged in a number of subsections, presents the results of tests of alternative models of bunching. The paper concludes with a summary of the findings and implications for the broader literature on bunching at policy thresholds.

2. Related Literature

This paper is related to two important strands of recent literature: (i) the role of third party reporting and withholding, and (ii) implications of thresholds, kinks and notches in a tax system.

As mentioned briefly in the introduction, the recent literature has underscored the importance of third party reporting in both developed and developing countries. An important contribution in this literature is the influential work by Saez, Kleven and Kreiner (2016) which shows that third party enforcement (including withholding at the source) is the critical institutional feature to understand the puzzle of high tax compliance in the face of low audit rates and punishments.

The effects of nonlinearities and discontinuities arising from government policies on economic choices of citizens have been the focus of a substantial literature, especially in labor and public economics. For example, Blundell and Hoynes (2004) analyze the effects of the requirement of a minimum of 16 hours work for UK family credit, and find clear evidence of bunching at exactly 16 hours of work. Closer to the subject of our analysis, there has been growing interests in understanding the role played by thresholds, kinks, and notches in the tax system. Slemrod (2016) and Kleven (2015) provide excellent reviews of this active area of research. In an interesting paper, Saez (2010) shows that the taxpayers in USA bunch at the first kink of the earned income tax credit (EITC) schedule, but there is no evidence of bunching at other kinks of EITC. He finds that the compensated elasticity of reported income with respect to the marginal tax rate is about zero for wage earners but substantially higher among those with self-employment income. Saez (2010) constructs an expected utility model of evasion with fixed costs of reporting and linear preferences which can explain these findings. Chetty et al. (2011) provide evidence of substantial bunching at a large kink of the Danish income tax schedule, but do not find any bunching at smaller kinks, consistent with Saez (2010). A related literature focuses on the tax notches (for example, change in average tax rates at a threshold); see, for example, the interesting analysis in the context of Pakistan by Kleven and Wasim (2013), and in the context of Costa Rica by Bachas and Soto (2017). Kleven (2015) provides an excellent treatment of the relative strengths of research designs that rely on kinks versus notches.

The implications of value added tax (VAT) exemption threshold have been analyzed in a number of recent papers. Among theoretical contributions, Keen and Mintz (2004) show that VAT thresholds create incentives for firms to bunch just below the threshold, and it distorts the firm size distribution in an economy. Onji (2009) reports empirical evidence in the context of VAT in Japan, by estimating the response of Japanese firms to a value-added tax threshold

that accorded preferential tax treatment to smaller firms (less than 500 million yen in sales). He finds evidence of firms bunching near the imposed threshold, suggesting that the large firms had incentives to "masquerade" as many small firms by separately incorporating business segments to avoid taxes.

Closer to the concerns of this paper, there is a small economic literature that analyzes possible interactions between withholding and evasion. Among the theoretical papers, Yaniv (1988) constructs a model to analyze how withholding of wage income by an employer can lead to evasion of the non-withheld taxes through non-filing of individual return, and Yaniv (1998) uses a prospect theoretic model to explain the evidence that firms that expect to get a refund are also more likely to submit a tax return. The empirical analysis of the effects of withholding has been a relatively neglected area in the literature. Using data from Costa Rica, Brockmeyer and Hernandez (2016) provide evidence that an increase in the withholding rate increases corporate tax revenue. In an interesting and careful analysis of withholding of personal income tax in Sweden, Engstrom et al. (2015) develop a regression kink and discontinuity design to provide credible evidence that individual tax payers with a preliminary deficit (tax liability greater than withheld amount) use deductions to escape from the loss domain. Using a prospect theoretic model with reference dependence, loss aversion, and linear preference, they provide an estimate of the loss aversion parameter. Linearity of preference seems a plausible assumption in their context, as they focus on small line item deductions. Their analysis does not explore the implications of linear preference for persistence in bunching, and the pattern of unbunching over time. In an important and insightful recent contribution, Rees-Jones (2018) presents a sophisticated analysis of bunching of individual income tax payers in USA and provides credible evidence of tax manipulations due to loss aversion. An interesting and important result in his analysis is that even with linear preference it may not be possible to isolate the loss aversion parameter from costs of manipulation. We provide complementary analysis by identifying other challenges in estimating the loss aversion parameter from the observed bunching at policy thresholds.

3. Theory: Models of Firm's Tax Choices in the Presence of Withholding

We begin with the standard Allingham-Sandmo model (henceforth A-S model) where with-

7

holding does not affect a firm's decisions because it does not affect the tax liability or the enforcement. Then we introduce a simple extension that incorporates the idea that claiming refund is costly. A second extension of the A-S model focuses on the possibility that claiming refunds usually increases the probability of an audit, *ceteris paribus*, and develops a model where the probability of audit jumps up discontinuously when a firm claims positive refund.[9] These three models are rooted in the expected utility theory of decision making under risk. The recent literature in tax enforcement and compliance has highlighted that the A-S model based on expected utility theory has some implications that are firmly contradicted by the evidence, and the prospect theory (Kahnemen and Tversky (1979)) provides a richer and more realistic framework. A fourth model developed below incorporates three major elements of prospect theory: reference dependence, loss aversion, and diminishing sensitivity (implying utility function is concave in the positive domain, but convex in the negative domain). In our context, the withheld amount provides a natural reference point.

3.1. Withholding in the Allingham-Sandmo (A-S) Model

Assume that a firm's tax liability is proportional to its profits. Let $\tau \in (0,1)$ be the profit tax rate. We define $\Pi_i = Y_i - C_i$ where Π_i is the profit of firm i, and Y_i and C_i are sales and costs respectively. So the true tax liability of firm i is $T_i = \tau \Pi_i$. The firm reports profit $\pi_i^* \leq \Pi_i$ and its tax payments are $t_i = \tau \pi_i^*$. When $\pi_i^* < \Pi_i$, the firm underreports its tax liability and evades taxes. When tax evasion is detected, the firm is forced to pay the evaded tax plus a penalty at a rate θ. We now introduce withholding by a third party. The third party withholds a proportion of the sales revenue of the firm and transmits it to the tax authority. Let the withholding rate be δ, so the amount withheld is δY_i. We define

$$\Psi_i = \frac{t_i}{Y_i} \quad ; \quad \tilde{\Psi}_i = \frac{T_i}{Y_i} \tag{1}$$

Thus $\tilde{\Psi}_i$ and Ψ_i are the true tax liability as a proportion of sales revenue and the actual tax payments as a proportion of sales revenue respectively, and a firm's true (no evasion) refund is

---

[9]While this is true in many developing countries, it is not necessarily true in developed countries such as Sweden where the refund is automatic, the taxpayer does not need to make any claim for it. Thanks to Per Engstrom for pointing this out. For a discussion on the refund process in Ecuador, please see section 5 below.

positive if $\tilde{\Psi}_i < \delta$, and the claimed refund is positive if $\Psi_i < \delta$.

Firm i chooses π_i to maximize its expected after-tax profit:

$$\max_{\pi_i} W_i = (1 - P) U_i (Z_i^g) + PU_i (Z_i^b) \tag{2}$$

$$where \quad Z_i^g = \Pi_i - \tau\pi_i$$

$$Z_i^b = (\Pi_i - \tau\pi_i) - \tau (1 + \theta) (\Pi_i - \pi_i)$$

and P is the probability of audit, Z_i^g is the after tax profit when firm is not audited, Z_i^b is the after tax profit when it is audited and has to pay penalty for tax evasion, and $\theta \in (0,1)$ is the penalty.[10] We assume that an audit reveals tax evasion with certainty. The utility function U_i is indexed by firm and captures possible risk heterogeneity among the firms. With a given probability of detection, the first order condition for an interior solution implies the following:

$$\frac{U_i' (Z_i^g)}{U_i' (Z_i^b)} = \frac{P\theta}{(1 - P)} \tag{3}$$

Proposition 1

Consider the optimal decision of a firm i with profits Π_i facing a tax enforcement regime $(P, \theta > 0, \delta > 0)$. Denote the optimal reported profit to the tax authority by π_i^*. There exists thresholds $1 > \hat{P} \geq \tilde{P}_i > 0$ such that:

(1.a) If $P \geq \hat{P}$, then $\pi_i^* = \Pi_i$; $\tag{4}$

(1.b) If $P < \tilde{P}_i$, then $\pi_i^* = 0$; $\tag{5}$

(1.c) If $P \in \left(\tilde{P}_i, \ \hat{P} \right)$, then $\begin{cases} \pi_i^* = 0 \ \text{if firm } i \text{ is risk neutral} \\ \pi_i^* \in (0, \ \Pi_i) \ \text{and} \ \dfrac{\partial \pi_i^*}{\partial P} > 0 \ \text{if firm } i \text{ is risk averse} \end{cases}$ $\tag{6}$

---

[10]The restriction that the penalty rate is less than 100 percent is motivated by the fact that we are not aware of any country where the fines for corporate tax evasion are as high as 100 percent. However, none of the results that follow depends on this assumption.

Proof:

Omitted. Please see the Online Appendix.

The important point in proposition (1) for our analysis is that a firm's optimal choices do not depend on the withholding rate δ, it only determines the amount of refund claimed once π_i^* is chosen according to proposition 1. Also, the results in proposition (1) are useful for understanding the effects of withholding in an audit trigger model (see section 3.3).

3.2. Transaction Costs in Claiming Refunds

A simple extension of the A-S model that allows for withholding to affect a firm's decision regarding tax payments can be due to transaction costs in claiming refunds, even though the level of profit declared is not affected by the withholding. Consider the case when $P \in \left(\tilde{P}_i, \ \hat{P} \right)$. Assuming that there is no additional costs associated with claiming the refund, then proposition (1) implies that a firm with $\Psi_i < \delta$ claims a refund even if it is small.[11] But if there are fixed costs of compliance in claiming a refund, then some firms might choose not to claim a refund. Let us denote the compliance costs of claiming the refund for firm i by $\kappa_i > 0$. Then the firm i with $\tilde{\Psi}_i < \delta$ will not claim any refund if the following holds:

$$\left(\delta - \frac{\tau \pi_i^*}{Y_i} \right) < \frac{\kappa_i}{Y_i} \Rightarrow \kappa_i > \left(\delta Y_i - \tau \pi_i^* \right) \tag{7}$$

One possibility for such a firm is to declare taxes at the withholding threshold, since it is not going to claim the refund anyway. We may thus observe some bunching of firms' declaring tax liability at (or around) the withholding threshold. However, note that if a firm declares taxes such that $\Psi_i < \delta$, and then does not claim the refund because of the costs of compliance, we should not observe the firms to bunch at the withholding threshold in the data where taxes are the reported taxes, not the amount withheld.

3.3. An Audit Trigger Model

In this section, we formalize the idea that when a firm declares taxes less than the withheld amount, the probability of audit increases. It may happen when the firms asks for positive refund, and the tax authority (SRI) asks for additional documentation. However, the probability of audit

---

[11]This is a good description of countries such as Sweden where the refund is automatic.

may be higher for such a firm even if it does not request for the refund (see the discussion in the context of Ecudaor in the next section). For simplicity, we assume that firms cannot hide the sales revenue, and thus tax evasion occurs through over-reporting of costs. Denote the reported costs by $C_i^d \in [C_i, Y_i]$. It is common that the tax authority uses the withholding rate as a focal point to allocate its auditing resources (confirmed by interviews with tax officials in Ecuador). If a firm declares a profit such that the implied (claimed) refund is positive, it triggers a discontinuously higher probability of audit. We assume that the probability of detecting tax evasion depends on the actual tax payments as a proportion of sales: $P(\Psi)$ is a negative function of Ψ_i, but it declines discontinuously when $\Psi_i \geq \delta$. Such an auditing scheme where the tax authority concentrates its resources on the firms below a threshold might be optimal as shown in the theoretical literature on optimal auditing (see Reinganum and Wilde (1985, 1986)).

The discontinuity in the $P(\Psi_i)$ function can induce the firms to bunch at or around the withholding threshold δ. To see this in a simple way, we consider a simple step function for $P(\Psi_i)$:

$$
P(\Psi_i) = \begin{cases} P_1 \text{ if } \Psi_i < \delta \\ P_2 \text{ if } \Psi_i \geq \delta \end{cases}
$$
$$
\text{with } \tilde{P} < P_2 < P_1 < \hat{P}
$$

The assumption that $P_1 < \hat{P}$ precludes the unrealistic case where the increase in probability in audit when a firm makes a positive refund claim is so large that it makes all of the firms declare taxes truthfully. We consider two types of firms separately depending on whether the true tax liability $\tilde{\Psi}_i$ is lower or higher than the withholding threshold δ. Consider the firms with true tax liability less than the withholding threshold (i.e., $\tilde{\Psi}_i < \delta$). When $P_1 < \hat{P}$, the risk neutral firms declare $\pi_i^* = 0$ and risk averse firms declare $\pi_i^* \in (0, \Pi_i)$.[12]

For our analysis, more relevant is the subset of firms with $\tilde{\Psi}_i \geq \delta$, the firms that face lower probability of audit P_2 if they declare tax liability truthfully, because their true refund is negative. If the probability of audit remains the same at P_2 across positive and negative refunds as is the

---

[12]Proposition 1 implies that they do not evade taxes and declare $\pi_i^* = \Pi_i$ if $P_1 > \hat{P}$.

case in the standard A-S model, then some firms may find it optimal to report costs $\left(C_i^d\right)$ high enough to make the refund positive. However, the discontinuous increase in the probability of audit in the AT model when a firm claims a positive refund may make it optimal to bunch at the withholding threshold. The set of firms that bunch at the withholding threshold and claim zero refund is identified in proposition (2).

Proposition 2

Assume that $\tilde{P} < P_2 < P_1 < \hat{P}$ and define $\mu_i \equiv \left[\frac{U_i'\left(Z_i^g\right)}{U_i'\left(Z_i^b\right)}\right]_{\pi_i \tau = \delta Y_i}$. Consider the firms with true tax liability higher than the withholding threshold, (i.e., $\tilde{\Psi}_i > \delta$). There exist two thresholds $\breve{\mu} > \tilde{\mu}$ such that the following results hold:

(2.a) The subset of firms that bunch at the withholding threshold belongs to the interval $\mu_i \in \left(\tilde{\mu}_i, \breve{\mu}_i\right)$.

(2.b) The relatively more risk averse firms, i.e., $\mu_i < \tilde{\mu}_i$, declare taxes higher than the withholding threshold, but less than their true tax liability, i.e., $\pi_i^* \in \left(\frac{\delta Y_i}{\tau}, \Pi_i\right)$.

Proof:

Omitted. Please see the online appendix.

Discussion

Note that μ_i is the ratio of marginal utility in good state (no audit) to that in bad state (with audit) evaluated at the reported taxes exactly equal to the withheld amount. The lower threshold $\tilde{\mu}_i$ is determined by the following condition:

$$\left.\frac{dW_i}{dC_i^d}\right|_{\pi_i = \frac{\delta Y_i}{\tau}} \geq 0 \implies \mu_i \geq \tilde{\mu}_i \equiv \frac{P_2 \theta}{(1 - P_2)}$$

Thus the lower threshold is such that firms with $\mu_i > \tilde{\mu}_i$ would prefer to declare taxes higher than the withheld amount. The upper bound is determined as follows:

$$\left.\frac{dW_i}{dC_i^d}\right|_{\pi_i = \frac{\delta Y_i}{\tau}} \leq 0 \implies \mu_i \leq \breve{\mu}_i \equiv \frac{P_1 \theta}{(1 - P_1)}$$

Thus the firms with $\mu_i < \breve{\mu}_i$ do not find it optimal to declare taxes more than the withheld

amount when facing the higher audit regime P_1. With $P_1 < \hat{P}$, the firms in the middle of the risk preference distribution bunch at the withholding threshold.

The magnitude of μ_i can be high due to two different factors: (i) low risk aversion, (ii) low profit Π_i. To see this transparently, consider the case where the utility function is of CRRA form:

$$U_i\left(Z_i\right) = \frac{Z_i^{1-\gamma_i}}{1-\gamma_i}$$

In this case, we have

$$\mu_i = \left[\left(\frac{Z_i^b}{Z_i^g}\right)^{\gamma_i}\right]_{\pi_i \tau = \delta Y_i}$$

It follows that $\dfrac{\partial \mu_i}{\partial \gamma_i} = \mu_i Ln\left(\dfrac{Z_i^b}{Z_i^g}\right) \leq 0$, because $\left(\dfrac{Z_i^b}{Z_i^g}\right) \in (0,1]$ in general. It is straightforward to check that $\dfrac{\partial \mu_i}{\partial \Pi_i} < 0$.

3.4. Withholding in A Prospect Theory Model: Do Reference Dependence and Loss Aversion Lead to Bunching at the Withholding Threshold?

An immediate and obvious implications of withholding a proportion of sales revenue of a firm is that it can act as a reference point when a firm is considering the cost-benefits of tax evasion. The fact that withholding threshold can act as a reference point for tax payers has been noted in the literature in the context of personal income tax evasion in Sweden and USA (see Engstrom et al. (2015), Rees-Jones (2017)). We thus take the withheld amount (δY_i) as the status quo at the time of tax decisions by a firm, instead of the total profit of the firm as is the case in the variants of Allingham-Sandmo models discussed above in sections (3.1) to (3.3). The utility function (denoted by V_i) is assumed to be of the following form:

$$V_i\left(X_i\right) = \begin{cases} U_i\left(X_i\right) & \text{if } X_i \geq 0 \\ -\lambda U_i\left(-X_i\right) & \text{if } X_i < 0 \end{cases}$$

where $\lambda > 1$ captures the strength of loss aversion, and $U_i\left(X_i\right)$ is strictly concave so that $V_i\left(X_i\right)$ is concave in the gains domain, and convex in the losses domain. The payoffs in good (no audit, denoted by superscript g) and bad (audit, denoted by superscript b) states are now

given as below:

$$X_i^g(\delta) = \delta Y_i - \tau \pi_i \, ; \, X_i^b(\delta) = \delta Y_i - \tau \pi_i - \tau(1+\theta)(\Pi_i - \pi_i)$$

The enforcement regime is same as in the basic Allingham-Sandmo model in section (3.1) above: the probability of audit is $P > 0$. We thus ignore the possibility that the firms may use a nonlinear transformation of the probabilities to weight the payoffs in different states of the world. This helps us to focus on the implications of reference dependence, loss aversion, and diminishing sensitivity.

The firm chooses $C_i^d \in [C_i, Y_i]$ to maximize the expected value of the prospect (suppressing δ from $X_i(\delta)$ for notational simplicity):

$$Max_{C_i^d} EW_i\left(X_i^b, X_i^g \mid P, \theta, \delta\right) = (1-P)V_i\left(X_i^g\right) + PV_i\left(X_i^b\right) \tag{8}$$

When the relevant C_i^d is such that the payoffs are in the gain domain in both audit and no-audit states, the optimal choices of the firm are exactly the same as the standard Allingham-Sandmo model based on expected utility theory. The most interesting cases for our analysis of possible bunching in tax payments are the ones where the payoff is positive in one state, but negative in the other. The intuition is that the kink in the utility function at zero when pay-offs span both the gain and loss domains may result in bunching in tax payments of firms. Thus our focus is on the case where the utility function has a kink in relevant domain of C_i^d. We treat the cases where the true refund (i.e., when firm declares $C_i^d = C_i$) is positive and negative separately.

First, consider the firms that would claim positive refunds when declaring costs (and thus taxes) truthfully, i.e., $\tilde{\Psi}_i < \delta$, and $X_i^g(\delta) = X_i^b(\delta) > 0$ when a firm declares $C_i^d = C_i$. This implies that payoffs in both states of the world remain positive, if the extent of over-reporting is small. An increase in reported costs C_i^d increases X_i^g (payoff when not audited) and decreases X_i^b (payoff when audited), thus X_i^b turns negative if the reported cost is high enough. It is easy to check, under certain restrictions on the magnitude of the penalty rate and the strength of loss

aversion, only the risk-neutral firms may bunch at $X_i^b(\delta) = 0$. The bunching is driven by the fact that the marginal cost of evasion function has a discontinuity at $X_i^b(\delta) = 0$. The important point for our analysis, however, is that we should not observe any bunching in *tax payments*, because the tax payments implied by $X_i^b(\delta) = 0$ are: $\tau\pi_i = \dfrac{1}{\theta}\left[\tau(1+\theta)(Y_i - C_i) - \delta Y_i\right]$. It is easy to check that the tax declared by such a firm is always lower than the withheld amount.

We now turn to the central case for our analysis where the refund is negative if the firm declares costs truthfully, i.e., $\tilde{\Psi}_i > \delta$. In this case, $X_i^g(\delta) = X_i^b(\delta) < 0$ if a firm declares truthfully, i.e., $C_i^d = C_i$. As C_i^d increases, $X_i^b(\delta)$ remains negative, while $X_i^g(\delta)$ moves towards zero, eventually becoming positive with a kink at $X_i^g(\delta) = 0$. If optimally located at the kink, the firm declares costs such that the tax payments exactly match the amount withheld. The marginal return function is discontinuous in this case, but the marginal cost of evasion is not. The expected marginal returns and costs are as follows (suppressing δ from X_i):

$$EMR = \begin{cases} -\lambda(1-P)\dfrac{\partial U_i(-X_i^g)}{\partial C_i^d} & \text{if } X_i^g < 0 \\[2mm] (1-P)\dfrac{\partial U_i(X_i^g)}{\partial C_i^d} & \text{if } X_i^g \geq 0 \end{cases} \quad ; \quad EMC = -\lambda P\dfrac{\partial U_i(-X_i^b)}{\partial C_i^d} \qquad (9)$$

Thus for a firm to optimally declare costs at the kink point (i.e., $X_i^g(\delta) = 0$) resulting in bunching of tax payments at around the withholding threshold, a firm's marginal cost of evasion has to fall inside the discontinuity in the marginal returns function noted above. To understand better the conditions under which we can expect bunching of tax payments at the withholding threshold, we consider two most widely used functional forms for $V_i(.)$ in the literature; the S-shaped power function that exhibits diminishing sensitivity and kinked linear preference. While the S-shaped power function has been most widely used as the value function in prospect theory, starting from the seminal work of Kahneman and Tversky (1979), many recent works on bunching at kinks and thresholds in tax systems assume linear preference (see, for example, Saez (2010), Engstrom et al. (2015)).[13]

---

[13]The model developed by Saez (2010) is not based on loss aversion, where fixed costs of reporting plays a critical role in an expected utility model of evasion.

The power value function is given as below:

$$U_i(X_i) = X_i^{\beta_i} \tag{10}$$

with $\beta_i \in (0,1)$ for a strictly concave function in the gains domain, and $\beta_i = 1$ for linear preference. As noted by Kahneman and Tversky (1979), preference homogeneity requires that the value function takes the power function form as in equation (10) above, and al-Nowaihi et al. (2008) show that loss aversion rules out different values for the exponent of the power function in the gains and losses domains (for a discussion, see Dhami (2016)).[14] So the S-shaped $V_i(.)$ function takes the following form:

$$V_i(X_i) = \begin{cases} X_i^{\beta_i} & \text{if } X_i \geq 0 \\ -\lambda(-X_i)^{\beta_i} & \text{if } X_i < 0 \end{cases}$$

The corresponding condition required for bunching of tax payments at the withholding threshold δ is:

$$\frac{\theta P}{1-P} < \left(\frac{\left|X_i^b(\delta)\right|}{\left|X_i^g(\delta)\right|}\right)^{1-\beta_i} < \frac{\lambda \theta P}{1-P} \tag{11}$$

It immediately follows that the second part of inequality condition (11) cannot be satisfied at $X_i^g(\delta) = 0$ if $\beta_i \in [0,1)$. Given a value of $\beta_i \in [0,1)$, the second part of inequality (11) is satisfied only if $\left|X_i^g(\delta)\right|$ is larger than a strictly positive threshold, thus ruling out not only exact bunching at the threshold, but also precluding bunching in a neighborhood $[\epsilon_-, \epsilon_+]$ of the threshold. This is intuitive as the power value function satisfies the Inada conditions.

The impossibility of bunching in the standard prospect theoretic model is due to a combination of two properties of the value function: homotheticity, and diminishing sensitivity ($\beta_i < 1$). One can generate bunching in this model by relaxing either of these assumptions, but the important question is whether the resulting value function is empirically grounded and theoretically consistent.

To see that a nonhomothetic value function can generate bunching, consider the simple case

---

[14]Preference homogeneity implies that when all prizes in a lottery are scaled up by a factor ϕ then the certainty equivalent of the lottery is also scaled up by ϕ.

where the value function is given as below (instead of equation (10) above):[15]

$$\tilde{U}_i\left(X_i\right) = \left(\Delta_i + X_i\right)^{\beta_i} \tag{12}$$

The inequality condition for bunching at the withholding threshold in (11) above now becomes:

$$\frac{\theta P}{1-P} < \left(\frac{\left|\Delta_i + X_i^b\left(\delta\right)\right|}{\left|\Delta_i + X_i^g\left(\delta\right)\right|}\right)^{1-\beta_i} < \frac{\lambda \theta P}{1-P} \tag{13}$$

It is clear that inequality (13) can be satisfied with $X_i^g\left(\delta\right) = 0$ as long as $\Delta_i \neq 0$, implying that bunching at the withholding threshold for a subset of the firms is possible. However, it is difficult, if not impossible, to find a plausible interpretation of the nonhomothetic term Δ_i, and we are not aware of any empirical foundation for such a value function in the behavioral economics literature.[16] When $\Delta_i < 0$, this specification effectively assumes that there is a second reference point in addition to the withholding threshold. This defeats the whole idea that the withholding threshold is the relevant reference point for a firm's tax decisions. We thus do not pursue this as a reasonable extension.

In contrast, there is a substantial literature in behavioral economics that provides theoretical and empirical justifications in favor of a linear (homothetic) value function (see, for example, Benartzi and Thaler (1995), Thaler et al. (1997), Nielson (2002), Dhami (2016)). As noted by Arrow (1971), linearity of preference (equivalent to risk neutrality in expected utility theory) may be a reasonable approximation for small stakes, and Rabin (2000) argues that it is also plausible for medium stakes. In the context of prospect theory, a kinked linear specification of $V_i\left(.\right)$ function has some additional desirable properties; it satisfies weak and strong loss aversion as defined by Nielson (2002), while the power value function does not. In the model above, when $\beta_i = 1$ (i.e., linear preference), it is possible for firms to optimally declare $C_i^d > C_i$ such that $X_i^g\left(\delta\right) = 0$. Given an enforcement regime (P, θ), bunching requires the following conditions: $P\theta < (1-P)$ and $\lambda > \dfrac{1-P}{\theta P}$. An immediate implication is that with weak tax administration

(low P and θ), it is easier to satisfy the first inequality, but it is also more likely to violate the second inequality given a degree of loss aversion λ. Thus in developing countries where tax enforcement capacity is low, it is less likely to observe bunching of taxpayers at the withholding threshold due to loss aversion and reference dependence. On the other hand, if P and θ are very high, then it is unlikely that the first inequality will be satisfied, which implies that bunching due to loss aversion is also less likely when enforcement regime is very strong. However, the recent evidence that individual taxpayers in Sweden and USA bunch in withholding threshold due to loss aversion presented by Engstrom et al. (2015) and Rees-Jones (2018) suggests that the first inequality is not violated in these cases. This is consistent with the widely held view that the enforcement regime in developed countries is not strong enough to account for the observed tax compliance in a an Allingham-Sandmo model of tax evasion (Saez et al. (2016)).

The result that with linear preference firms may bunch their tax payments at the withholding threshold due to loss aversion is consistent with Engstrom at al. (2015) where the focus is on individual taxpayers, but we establish it in a more general model with an enforcement regime. The result that with S-shaped power utility function exhibiting diminishing sensitivity, firms cannot bunch at the withholding threshold has not been noted in the literature, to our knowledge. Perhaps more important, as we discuss below, the assumption of kinked linear preference generates a set of testable predictions (please see section (4) below), which are not considered in the current literature, to our knowledge.

Proposition 3

(3.a) Assume that the firms have S-shaped power utility function $V_i(.)$ that exhibits loss aversion, diminishing sensitivity (concave in the gains domain, and convex in the losses domain, i.e., $\beta_i \in [0,1)$),) and reference dependence with the withheld amount as the reference point. Then the tax payments of the firms do not bunch at the withholding threshold irrespective of whether a firm expects a positive $\left(\tilde{\Psi}_i < \delta\right)$ or negative $\left(\tilde{\Psi}_i > \delta\right)$ refund when declaring taxes truthfully.

(3.b) Assume that the utility function is linear in both gains and losses domains, i.e., $\beta_i = 1$. Consider the firms that expect a negative refund when declaring taxes truthfully, i.e., with $\tilde{\Psi}_i > \delta$. Then the tax payments of firms bunch at the withholding threshold if the loss aversion is high

enough so that $\lambda > \dfrac{1-P}{\theta P}$, *and the penalty for evasion is low enough so that* $\theta < \dfrac{1-P}{P}$.

(3.c) The tax payments of the firms with $\beta_i = 1$ *and* $\tilde{\Psi}_i < \delta$ *do not bunch, and they declare tax payments to the left of the withholding threshold.*

4. Empirical Strategy

4.1 Identification Problem: Behavioral Response to the Threshold or Lump in the Firm Distribution?

The analysis above shows that bunching around the withholding threshold can arise from optimizing choices of firms trying to evade taxes facing a withholding and enforcement regime. There is, however, another possibility that can give rise to bunching of firms around the threshold which has little to do with a firm's optimizing behavior. The observed bunching could simply reflect a mass in the true distribution of firm's profit/sales ratio determined by factors such as preference and technology. While firms' profit/sales ratios are determined by many characteristics including entrepreneurial ability, we can illustrate the possibility by using the simple model above where firm's differ in terms of risk preference only. Consider the AT model where bunching is due to evasion facing a change in the audit rate across the withholding threshold. According to the audit trigger model, the mass of bunched firms is given by

$$B\left(\delta; P_1, P_2, \theta\right) = \int_{\tilde{\mu}}^{\breve{\mu}} dF(\mu_i)$$

where $F(\mu_i)$ is the CDF of the firms with true tax liability higher than the withholding threshold. One can observe a lump in the distribution of firms' profit tax payments around δ in the absence of any bunching induced by withholding if the following conditions hold: (i) there is a lump in the risk heterogeneity below $\breve{\mu}$, (ii) firms with similar risk preference also have similar profits and sales revenue.

While the withholding rate may coincide with such a lump in the distribution by pure chance, it can also be outcome of choices by the tax administration. If the tax administration has a good estimate of the true distribution of profits and sales, they may set the "right" withholding rate–one that facilitates tax payer's compliance. In these cases, the observed bunching around the withholding rate may not indicate any behavioral response by the firms to the threshold.

This is the central identification problem for our empirical analysis: how do we know that the observed bunching is not due to a lump in the firm distribution due to lumps in the distribution of risk preference and technology, but is an outcome of firm's optimal choices as illustrated in the models above? It is extremely difficult, if not impossible, to reject this alternative hypothesis in a cross-section survey, or in panel data where the withholding rate does not vary across firms and over time. An important advantage of our study is that we can exploit quasi experimental variations in the withholding rate across sectors and over time to provide credible evidence on the relevance of optimizing choices of firms in the observed bunching. To solve the identification problem we exploit a simple insight: if the bunching is due to technological or other factors unrelated to withholding, firms will stay bunched at the same threshold when the withholding rate is changed by a policy reform.

Testable Prediction 1

(TP 1) If the bunching observed at around the withholding threshold is due to factors unrelated to withholding such as technology, then when the withholding rate changes, the bunching should remain largely undisturbed. If the alternative hypothesis that bunching is due to optimizing choices by the firms facing the withholding threshold, then the bunching in tax payments at around the initial threshold would not be observed after the change in the withholding threshold, the distribution would become smooth.

4.2 Empirical Strategy to Discriminate Among Alternative Models

4.2.1. Testing the Null Hypothesis of Bunching Due to Transaction costs

We develop a test of the null hypothesis that bunching is due to transaction costs by exploiting a quasi-experimental policy reform in Ecuador in June 2007 that increased the withholding rate from 1 percent to 2 percent in some of the industries. The policy reform can be modeled as an increase in the withholding rate from δ to 2δ. We focus on the implications of the increase in the withholding rate on the tax payments of the firms in 2007 that bunched at the δ-threshold in the earlier years. We emphasize here that the new withholding rate was imposed in the middle of the year, so we should not expect any bunching of tax payments at the new threshold in 2007.[17]

---

[17]A related point is that bunching at the new threshold for evasion is likely to take time as the relevant firms adjust their accounting to the new threshold.

When the threshold is δ, the set of firms that bunch because of transaction costs is given by $\{i\} \ni \tilde{\Psi}_i < \delta$ and $\kappa_i > (\delta Y_i - \tau \pi_i^*)$. Thus the firms that bunch at the threshold are those with relatively less (true) tax liability and high transactions costs. Facing the higher withholding rate 2δ, many of these firms may find it worthwhile to claim the refund, as the change in the withholding rate should not affect the transaction costs. The set of firms that bunched at the δ threshold in 2006, but declare lower taxes after the reform and claim a positive refund is given by the following conditions: $\{i\} \ni \tilde{\Psi}_i < \delta$ and $(2\delta Y_i - \tau \pi_i^*)_{2007} > \kappa_i > (\delta Y_i - \tau \pi_i^*)_{2006}$. This implies that if bunching is driven by transactions costs, we should observe a decline in the tax declarations in 2007 of many of the firms that bunched at around the δ-threshold in 2006. Note that the 100 percent increase in the withholding rate makes this test credible, the effects of the policy change on the withheld amount can be canceled out only in the implausible case where the sales of a firm falls by 50 percent from 2006 to 2007. Note that even if a firm does not declare lower tax liability when the withholding rate is doubled, they would never declare higher taxes if transaction costs are the driving force behind the observed bunching. In the empirical implementation, we use a difference-in-difference design with firm fixed effects to take into account the changes in profit from 2006 to 2007. Since our analysis focuses on the profit/sales ratio, the changes in sales volume due to demand shifts cannot be responsible for the results.

The implications of an increase in the withholding rate in the LPT and AT models are different: they do not preclude the possibility that some of the firms will declare higher taxes in 2007 following the reform. First consider the AT model: the policy reform affects the tax payments of the firms with $\tilde{\Psi}_i \in (\delta, 2\delta)$ in an audit trigger model, because the audit probability faced by them increases from P_2 to P_1 even if they declare the tax liability truthfully.[18] Note that these firms will not bunch at the new threshold because only firms with $\tilde{\Psi}_i > 2\delta$ can bunch at the new withholding threshold. Many of the firms that evaded taxes and bunched at the δ-threshold in 2006 will declare higher taxes facing the higher probability of audit P_1 in 2007 because for any risk averse firm i , we have $\dfrac{\partial \pi_i^*}{\partial P} > 0$ according to proposition (1.c) above.

---

[18]For all other firms, the audit probability remains the same as before if they declare their tax liability truthfully, but the threshold where the audit probability changes is different at $2\delta Y$.

In the LPT model, an increase in the withholding rate affects the firms with true tax liability in the interval $\tilde{\Psi}_i \in (\delta, 2\delta)$ because they get a positive refund after the increase in the withholding rate when reporting costs truthfully, but owed taxes in the initial withholding regime (δ-regime). With linear preference firms can bunch because of loss aversion only if they get negative refund when reporting truthfully, the firms that bunched at the δ-threshold to evade taxes previously will not bunch at the new withholding threshold 2δ, but some of them may bunch at $X_i^b(2\delta) = 0$. As noted before, this implies that the tax payments do not bunch at the withholding threshold δ or 2δ. However, whether a firm declares higher tax liability depends on the profitability of a firm. It is easy to check that with the higher withholding rate, the tax payments implied by $X_i^b(2\delta) = 0$ are higher than tax payments of the firms bunched at δ (i.e., δY_i) when the following condition holds: $\left[\dfrac{Y_i - C_i}{Y_i} \right] > \left(\dfrac{\delta}{\tau} \right) \left(\dfrac{1 + \theta}{2 + \theta} \right)$. So after the increase in the withholding rate, some of the firms with higher profit that initially bunched at the δ-threshold and paid taxes equal to δY_i will declare higher tax payments, while others may declare lower tax payments after the withholding rate increases to 2δ. The response of tax payments by the firms that bunch at the δ-threshold before 2007 thus cannot discriminate between the AT and LPT models for which we develop different tests in the following section.

Testable Prediction 2

(TP 2) If the bunching at the δ-threshold is driven by transaction costs in claiming refunds, some of the initially-bunched firms would reduce their tax payments, and no firm would increase its tax payments, following an increase in the withholding rate, ceteris paribus. In contrast, if the bunching is driven by the discontinuity in the audit regime or loss aversion in a LPT model, the tax payments by some of the initially-bunched firms would increase after an increase in the withholding rate.

4.2.2. Bunching and Tax Evasion: Kinked Linear Preference or Discontinuity in the Audit Regime?

The theoretical results in propositions (2) and (3) in section 3 suggest a simple strategy to discriminate between the LPT and AT models: use indicators of risk heterogeneity across firms, and test the null hypothesis implied by LPT that the probability of bunching is the highest for

the firms most likely to be risk-neutral (linear preference). There is a substantial theoretical and empirical literature that suggests that the large firms are less risk averse, because they can diversify the risk, and also because less risk-averse managers/owners are more likely to self-select into large firms (Arrow (1993), Greenwald and Stiglitz (1990), Engel, Fischer, and Galetovic (1998), Herranz et al. (2015), Greenwald and Stiglitz (2014)). The more risk averse firms rely on self-financing for investment, and consequently accumulate less assets, and do not grow large.[19] Thus firm size measured by assets can be used as a plausible indicator of risk heterogeneity among firms; the larger a firm is, the more likely it is to be risk-neutral. In contrast, if the AT model is appropriate where bunching is driven by a discontinuity in the audit regime around the threshold, the relation between the firm-size and probability of bunching is inverted U-shaped, because only the firms in the middle of the asset size distribution are likely to bunch as they are most likely to satisfy the condition that $\mu_i \in (\tilde{\mu}, \breve{\mu})$.

The above tests of the AT and LPT models based on firm size are valid if loss aversion does not vary significantly across firm size. The available evidence supports this assumption (see Kapoor (2017)).[20] One can argue that the large firms are also likely to be less loss-averse, and thus not likely to bunch according to the LPT model. However, this also implies that the medium (and small) firms will also not bunch because they are likely to be not only more loss averse, but also more risk averse ($\beta_i < 1$).

A second and more definite test of LPT relies on the observation that $\beta_i = 1$ implies that the expected marginal returns to and marginal costs of evasion do not depend on the profit of a firm; implying that changes in a firm's revenue and costs cannot change a firm's bunching status as long as P, θ, and δ are unchanged. This has important implications for persistence in bunching over time. If a firm bunches at the withholding threshold in year t, it will bunch in all other years when P, θ, and δ remain the same, because the changes in revenue and costs over the years do not affect the bunching decision so long as true tax liability remains higher than the

---

[19]The available evidence shows that the small firms use less debt and buy less insurance (Herranz et al. (2015), Collier et al. (2016)). For evidence that large firms are more able to withstand negative shocks by using their better access to credit market, see Khwaja and Mian (2008). As noted by Arrow (1993), large firms may also have more stable internal capital supply.

[20]Using experimental evidence from restaurant industry, Kapoor (2017) shows that the effects of risk preference on loss aversion is not statistically significantly different from zero. See Table 3 in Kapoor (2017).

withheld amount, i.e., $\tilde{\Psi}_i > \delta$. When the firms experience shocks to revenue and costs in a way to make its true tax refunds switch from negative to positive, then they will unbunch from the withholding threshold, but bunch at the threshold where $X_i^b(\delta) = 0$ according to proposition (3) above. As shown in proposition (3.c) above, such firms always declare taxes *lower* than the withheld amount.

In contrast, the audit trigger model shows that firms in the middle of the μ distribution bunch, i.e., firm i bunches if $\mu_i \in (\tilde{\mu}, \breve{\mu})$, assuming that $P_1 < \hat{P}$ which is a plausible assumption in most, if not all, countries, as noted earlier. Since $\dfrac{\partial \mu_i}{\partial C_i} > 0$, and $\dfrac{\partial \mu_i}{\partial Y_i} < 0$, it follows that if there are significant changes in revenue and/or costs, a firm can bunch in one year, but not in another even though its true tax liability is higher than the withheld amount in all years, and the enforcement and withholding regime has not changed. We summarize the above discussion in the following testable predictions.

Testable Predictions 3

Assume that P, θ, and δ remain unchanged over the relevant years under consideration.

(TP 3.1) Since the large firms are more likely to be risk neutral (linear preference), the propensity to bunch will be highest for the firms that belong to the right tail of firm size distribution under the LPT model. Under the alternative hypothesis of AT model, the firms that fall in the middle of the farm size distribution will be most likely to bunch, making the relation between farm size and probability of bunching inverted U-shaped.

(TP 3.2) Under the null hypothesis that bunching is generated by linear preference with a kink due to loss aversion (LPT model), a firm that bunches at the withholding threshold at period t, will also bunch at other periods irrespective of the changes in its sales revenue and costs over the years as long as $\tilde{\Psi}_{it} > \delta$,$\forall t$. If the revenue and costs of a firm that bunched at the withholding threshold in year t change in year $t+1$ in a way so that it switches from $\tilde{\Psi}_{it} > \delta$ to $\tilde{\Psi}_{i(t+1)} < \delta$, then the firm unbunches and declares tax liability smaller than the withheld amount in year $t+1$.

(TP 3.3) Under the alternative hypothesis that bunching at the withholding threshold is due to the audit trigger model (AT model), some firms that bunch at period t, may not bunch at the withholding threshold in other years, even though $\tilde{\Psi}_i > \delta$ in all of the years. Under the AT model, some of the firms that unbunch may declare taxes higher than the withheld amount.

5. Tax Administration and Withholding System in Ecuador

In this section, we provide some details about the tax administration in Ecuador (SRI) and the withholding mechanism. Ecuadorian firms are taxed on their profits using a flat tax rate of 25%.[21] Firms are required to file year t tax returns between February and April of year $t+1$. The withholding system is a mechanism where companies designated by the SRI as withholding agents are required to deduct and withhold a fixed percentage of the payments they make to other firms. This fixed percentage is known as the withholding rate. Every month, withholding agents must report and transfer all withholdings to the tax authority. Firms can deduct their withheld funds, including those from previous years (up to five years), from their current tax liability. Sales to final consumers are not subject to withholding.

When the withheld amount exceeds the firm's tax liabilities, the SRI does not automatically grant a refund. The SRI keeps the difference between the withheld funds and the firm's tax liability unless a petition for a refund is filed. However, if the withheld amount exceed tax obligations, an Ecuadorian firm can request a tax refund due to over payment and/or request that the positive outstanding balance be used as a credit for the future (the credit has to be used within three years after the request is approved). These requests have to be approved by the tax administration and strictly adhere to procedures described in the Ecuadorian Tax Code's (Article 119).

Firms applying for a refund must submit a refund application as well as updated credentials of their legal representative, bank statements, invoices, and other documents that support the claim. Once a request is submitted, the SRI verifies the validity of the request and its documentation. During this process cross-checks are performed using third party data. If the tax administration detects irregularities, it can trigger an audit or, in more serious cases, suspend the refund process and start a "complementary procedure" aimed at reassessing the firm's tax liability (Ecuadorian Tax Code Article 131). This action can result in penalties according to the severity of the infringement. Thus requesting a tax refund increases significantly the intensity at which a firm is monitored by the SRI.

---

[21]This tax rate was unchanged until 2010. The common and unchanged tax rate helps us identify the effects of withholding threshold.

SRI uses data from many sources to compute a tax-evasion-risk index for every tax payer (firms and individuals) in Ecuador which is used to prioritize auditing efforts. Specific details about the construction of these indexes are confidential. It should be noted, however, that in addition to discrepancies between third-party and self-reported entries in the tax return, deadline noncompliance, any differences between the withheld amount and tax liability affect the index. In particular, when a firm's withheld taxes exceed the tax obligations, audit risk can automatically increase even if the firm does not request a refund.

The 1988 Tax Law set a maximum of 10 percent upper bound on the withholding rates. In 1989, a Ministerial Decree set the 1% withholding rate which for most goods and services remained unchanged until 2007. Withholding rates can be unilaterally changed by the tax authority without any approval of the legislative or executive power (assuming a rate not higher than the maximum of 10 percent). These rates in a given period vary according to the goods or services being purchased, and also have changed over time. Table 1 shows how withholding rates for the four categories of products we study have changed since 2003. In July 1, 2007, withholding rates for sales of retail goods and other services (including construction) increased from 1 to 2 percent. This was an important change that affected a large share of commercial transactions. But the withholding rate for sales of transportation services remained constant at 1 percent over this period; it was not affected by changes in withholding rates in 2007. While there are no official reports explaining why this was the case, transport sector is a highly organized group with unions operating at regional and national levels. It is likely that the government exchanged this exemption for political support. The important point for us is that the exemption was determined by political calculus, and thus does not reflect any systematic bias with respect to the tax liability of a firm.

The withholding rates listed on Table 1 apply to *products* rather than to *firms*. For example, a firm that sells both manufacturing goods and transportation services could be subject to different withholding rates after the policy change in July 1, 2007. However, such cases of a single firm spanning multiple withholding rates are highly unlikely. While there is no one to one correspondence between the firm's ISIC economic activity and the products it sells, it is reasonable to assume that a) all firms in the manufacturing and the retail sector are subject to

the withholding rates that apply to tangible goods purchases, b) all firms in the construction sector are subject to the withholding rates that apply to real estate construction activities, and c) all firms in the transportation sector are subject to the withholding rates that apply to private passenger transport and public and private freight services.

6. Does Withholding Lead to Bunching Due to Firm's Optimization?

In this section, we test if Ecuadorian firms' tax/sales ratio (ie., Ψ_i) concentrates around the withholding rate ($\delta = 0.01$) before 2007 reform. We start by describing the data, and then show graphical and econometric evidence that firms do bunch around the withholding threshold and that the bunching is due to firm's optimizing choices facing the withholding and enforcement regime.

6.1. Data

Our empirical analysis is based on administrative data from the Ecuadorian Tax Authority (SRI). We obtained the universe of corporate profit tax returns from for years 2004, 2005, 2006 and 2007. These returns are reported on form F101 and are submitted once a year. The fiscal year coincides with the calendar year, and returns are due in April of the following year.

We focus on the firms in the following four economic activities: construction, manufacturing, wholesale trade and retail, and transportation. This allows us to exploit the variation in the withholding rates both across industries and over time described in the previous section. After eliminating duplicates and observations with missing data, we are left with an unbalanced panel of about 12,000 firms per year.[22] We thus have three years of panel data from 2004 to 2006 during which the withholding and enforcement regime did not change. These three-year panel data allow us to implement the testable predictions (3) which help discriminate between the LPT and AT models. The policy reform in 2007 that increased the withholding rate from 1 percent of sales to 2 percent helps us discriminate between the transaction costs model against the LPT and AT models.

Details about the variables and descriptive statistics for the years 2006 and 2007 are provided in Table 2. Ecuadorian firms' assets in these industries averaged $1.7 and $1.9 million in 2006

---

[22]We focus on firms that report positive sales.

and 2007, respectively. Their mean annual revenue (sales) increased from $2.9 million in 2006 to $3.3 million in 2007. Average profit tax liability also increased from $28,000 to $35,000.

6.2. Bunching of Tax Payments at the Withholding Threshold : Graphical and Econometric Evidence

We first provide some graphical evidence on bunching by estimating the distribution of the firm's tax/sales ratio (Ψ_i) in each of the four industries in our data set. However, even if all firms in our sample report a tax liability that is exactly equal to their withheld funds, we do not expect to see a degenerate distribution of the tax/sales ratio for two reasons. First, the sales made to final consumers are not subject to withholding, and some firms in our sample sell some of their products to consumers. Second, the firms can use the withheld funds from previous periods (up to five years) to pay current year tax liabilities. For these reasons, even if all firms in our data declare taxes that are identical to their withheld funds, the distribution of the reported tax/sales ratio should be concentrated around the withholding requirement.

Figures 1 to 4 show histograms of the tax/sales ratio by year for firms in the construction, manufacturing, retail and transport sectors, respectively. Between 2004 and 2006, there is a noticeable clustering of firms around the 0.01 bin in all of the industries, which coincides with the withholding requirement of 1%.

To formally test for bunching patterns in the data we follow the approach suggested by Saez (2010). That is, we say that a firm "bunches" if its reported tax/sales ratio lies within a small distance (ρ) from the withholding rate. For robustness, we choose three alternative values for ρ (0.08%, 0.09% and 0.10%). The variable B_{ai} equals one if firm i's tax/sales ratio in 2006 is between 0.0092 and 0.0108. Similarly, the variables B_{bi} and B_{ci} equal one if the 2006 tax/sales ratio falls within the intervals [0.0091, 0.0109] and [0.0090, 0.0110], respectively. Using these definitions, we find that about 6%, 8% and 10% of firms in our sample "bunch" around the withholding rate, respectively.

For each of the three bands defined above, two symmetric surrounding intervals are computed: [1%-2ρ, 1%-ρ] and [1%+ρ, 1%+2ρ]. To estimate if there is excess bunching, we simply estimate the difference between the fraction of firms in the interval around the withholding rate and the fraction of firms in the two surrounding bands. Let H^* be the share of firms in the bunching

interval around the withholding rate and H^+ and H^- be the share of firms in the upper and lower surrounding bands, respectively. As in Saez (2010), the share of firms in each of the three bands is estimated by simultaneously regressing a dummy variable for belonging to each interval on a constant in the sample of firms belonging to any of these intervals. Excess bunching is defined as $E = H^* - (H^- + H^+)$. For each of the three interval definitions, we estimate E and test if it is statistically different from zero.

The first panel of Table 3 reports the test's p-values for each industry using the 2004-2006 pooled sample. Notice that for all values of ρ we can reject the null of no-bunching at virtually any standard significance levels. The statistical tests confirm the existence of excess mass of firms reporting a tax/sales ratio close to 1%. Tests are also performed using the 2006 sample only. Results shown on the second panel of Table 3 confirm that during 2006 firms in the four industries we consider bunch near the withholding rate.[23]

6.3. Evidence on the Effects of the Change in the Withholding Rate in 2007

We construct histograms of tax/sales ratio for the year 2007 when withholding requirements of some industries increased from 1% to 2%. A comparison of the 2007 histograms with the corresponding histograms for 2004, 2005, 2006 in figures 1-4 reveals interesting patterns. The most striking is the fact that the spike at the 0.01 tax/sales ratio bin observed in 2004-2006 completely disappears in 2007 in those industries where withholding rates were modified (construction, manufacturing and trade).[24] Moreover, the spike in the density of the tax/sales ratio for firms in the transportation sector seems to remain constant at 0.01 (the unchanged withholding rate). We also formally test if the 2007 tax/sales ratio densities display excessive bunching at 1%. Results displayed in the third panel of Table 3 suggest that there is no evidence of bunching except for the transport industry where the withholding rate did not change. These results contradict the interpretation that the bunching in 2004- 2006 at around the withholding threshold is due to underlying heterogeneity in firm characteristics, and are consistent with the

---

[23]We also use an alternative method due to Kleven et al. (2015) to compute excess bunching around the withholding threshold and obtained similar results. The details are available from the authors.

[24]Note that the change in withholding rates occurred in the middle of the year (June 2007) and affected only the second half of the year sales. Thus, one cannot expect to see bunching patterns around the new withholding rate of 2%.

predictions from our theoretical model.

7. Evidence on Alternative Models of Bunching
7.1. Is Bunching at the Withholding Threshold Due to Transactions Costs?

To test if the bunching observed in the years before the change in withholding rate in 2007 is consistent with the transaction costs model, we test the prediction that the firms that bunch at the withholding threshold in 2006 are likely to reduce their tax payments in 2007, and none of them can increase their tax payments. To ensure that the estimated effects of the change in withholding rate is not driven by changes in profitability from 2006 to 2007, we implement a difference-in-difference-in difference (or triple difference) strategy with firm fixed effect where the dependent variable is the difference between the 2007 reported profit tax to sales ratio and the 2006 counterpart. The independent variable of interest B is a binary indicator that takes the value of 1 if a firm's tax payments bunched at around the withholding threshold in 2006, i.e., the tax/sales ratio was close to $\delta = 0.01$. The Results of this test are shown in Table 4. The first column reports the estimates for the firms in industries that experienced an increase in the withholding rate (manufacturing, construction and retail) in 2007 while the sample used for the second column consists of the firms that experienced no change in withholding rates (transport sector). The estimates in each column provides the difference-in-difference (DID) estimates using the firms that do not bunch as the comparison. When we subtract the DID estimate in the second column (assuming the estimate is statistically significant) from that in the firms column, we get the triple-difference (DIDID) estimate of the effects of the change in the withholding rate on the tax declarations of the firms that bunched in 2006. When the estimates in the second column are not statistically different from zero, then the triple difference estimates are the same as the DID estimates.

The results provide evidence that the set of firms that bunched at around the threshold in 2006 shows a significantly larger increase in reported profit rates in 2007 only when withholding rates increased from 1 percent to 2 percent. For example, the coefficient of interest in the first row and first column (0.001) suggests that firms that bunched in 2006 in the manufacturing retail and construction sectors increased their reported profit-rates by 0.001 more, (about 10% more than the average profit rate at baseline) than their counterparts that did not bunch. The

estimate is statistically significant at the 1% significance level and economically important. On the other hand, when we analyze an industry that experienced no changes in δ in 2007, we find no differences: the coefficients of interest in the second column of Table 4 is a precisely estimated zero. Thus the triple difference estimate is in fact equal to the DID estimate reported in column 1. The estimates for other definitions of bunching in rows 2 and 3 are similar. They show that the coefficient is statistically significant at the 5 percent only for those firms that experienced an increase in the withholding rate. The estimates for the firms that did not face a higher withholding rate in 2007 are statistically not different from zero. The evidence that the DID estimates for the firms in the placebo industry (without the policy change) are not statistically significant thus provides support the the validity of the DID estimates.

The evidence in Table 4 thus rejects the null hypothesis that the mechanism behind the observed bunching of tax payments in 2006 is transaction costs in claiming refunds. However, as noted earlier, the evidence that the firms that bunched in 2006 increase their tax declaration in 2007 is consistent with both the LPT and AT models.

7.2. Linear Prospect Theory (LPT) vs. Audit Trigger Model (AT)

In this section we discuss evidence on the testable predictions (TP.3) that help us discriminate between the LPT and AT models.

The Relationship Between a Firm Size and Probability of Bunching

According to the testable prediction TP.3, only the risk-neutral (linear preference) firms bunch under the null hypothesis that LPT is the correct model of tax evasion by the firms. Moreover, the relation between risk averseness and the probability of bunching is inverted-U shaped under the AT model. We use the rank of a firm in the distribution of asset-size as an indicator of risk preference and risk bearing capacity of a firm with the firms at the right tail expected to behave as approximately risk-neutral.

Table 5A reports estimates of the partial correlation between indicators of of large farm size and the probability of bunching for three different definitions of bunching. We use four different definitions of "large firm", corresponding to firms that belong to more than 75th, 90th, 95th, and 99th percentile of asset size distribution. The estimates are consistently negative and

statistically significant at the 1 percent level. The evidence thus shows that the large firms are less likely to bunch which can be interpreted as suggestive evidence against the prediction from the LPT model that only the risk-neutral firms bunch at the withholding threshold.

Figure 5 shows the percentage of firms that bunch by asset decile; the relation seems inverted-U for all three definitions of bunching. Table 5B reports the results from formal test of the null hypothesis that the relation between the probability of bunching and firm size is inverted U. The standard approach to testing a U or inverted-U relation is to include a quadratic term and check the sign and significance of the coefficient on the quadratic term. However, as noted by Lind and Mehlum (2010), such a test cannot discriminate between a true inverted-U and a monotonically increasing concave function. So in addition to estimates from a quadratic specification, we implement the Sasabuchi t-test and use the Fieller method to estimate the implied extremum point and its 95% confidence interval as developed in Lind and Mehlum (2010). Consider the following quadratic regression specification:

$$B_{it}\left(\delta = 0.01\right) = \theta_0 + \theta_1 F_{i(t-1)} + \theta_2 F_{i(t-1)}^2 + \epsilon_i \tag{14}$$

where $B_{it}\left(\delta = 0.01\right)$ is a dummy variable that takes on the value of 1 when the firm i bunches at the 1 percent withholding threshold and $F_{i(t-1)}$ is the measure of firm size in terms of asset rank in period (t-1). The existence of an inverted-U can be tested on an interval $\left[F_i^l, F_i^h\right]$ with the following combined null hypotheses against the combined alternatives:

$$
\begin{aligned}
H_0 &: \theta_1 + 2\theta_2 F_{i(t-1)}^l \leq 0; \quad \theta_1 + 2\theta_2 F_{i(t-1)}^h \geq 0 \\
H_A &: \theta_1 + 2\theta_2 F_{i(t-1)}^l > 0; \quad \theta_1 + 2\theta_2 F_{i(t-1)}^h < 0
\end{aligned}
\tag{15}
$$

The estimates from the regression equation (12) above with t=2006 and $t-1$=2005 are reported in the top panel of Table 5B. The coefficient on the quadratic terms is negative and statistically significant at the 1 percent level, while the coefficient on the level term is positive and statistically significant at the 1 percent level. The evidence thus is consistent with a monotonically increasing concave relation or an inverted-U.

The lower panel of Table 5B reports the results of the test of the null hypothesis that the

relation is either convex or monotonically increasing concave function against the sharp alternative that the relation is inverted-U. The test of inverted-U in inequalities (13) is implemented using Sasabuchi (1980) t test, and the results are in row 5 of Table 5B. In addition, we estimate the implied extremum point and its 95% confidence interval using the Fieller method. The estimates reject the joint null that the relation is either monotone or U-shaped. The evidence thus provides strong support to the hypothesis that the relation between probability of bunching at the withholding threshold and assets of a firm is inverted-U shaped as predicted by the AT model of bunching.

Intertemporal Persistence in Bunching and the Pattern of Unbunching

The evidence on the relation between firm size (measured in terms of assets) and the probability of bunching is suggestive of firms behaving more according to the audit trigger model. However, firm size is an imperfect proxy for risk preference and risk bearing capacity of a firm.

In this section, we report more definitive tests in terms of persistence in bunching and the pattern of unbunching. A simple but convincing way to test whether there is strong intertemporal persistence in bunching before the change in withholding rate in 2007 is to check the proportion of firms that bunch in all three years (2004-2006) conditional on bunching at least once. The estimates for three different definitions of bunching are 3.2 percent, 3.6 percent, and 4.5 percent respectively (see row 1 in Table 6). This is strong evidence against the hypothesis that if a firm bunches in one year, it should also bunch in the other years if the enforcement and withholding regime is unchanged, and raises doubts about the LPT model as a description of the central tendency in the data.[25]

However, as noted in testable prediction (TP.3.2), a firm that bunched in year t can unbunch in the year $t+1$ under the LPT model when it experiences shocks such that its true refund switches from negative (in t) to positive (in $t+1$). The testable implication of LPT model in such cases of unbunching is that they will declare taxes less than the withheld amount in the year they unbunch. Rows 3 and 5 in Table 6 show the proportion of firms that declare more than

---

[25]One can argue that linearity will only be approximately satisfied in the data even if $\beta_i = 1$, and we should observe some unbunching. However, the fact that more than 95 percent of the firms unbunch in a span of only 3 years suggests that significant curvature is required in the value function to rationalize this.

the withheld amount when they unbunch in the year 2005 and 2006 respectively conditional on bunching in the year before. Given a definition of bunching, we classify a firm as unbunched when its declared tax/sales ratio in year $t+1$ falls outside the band of bunching, but the declared tax/sales ratio in the year t was inside the band. The estimates in row 3 of Table 6 refer to the firms that bunched in 2004, but unbunched from the withholding threshold in 2005, and 35 percent of the unbunched firms declare taxes higher than the withheld amount. The evidence in row 5 of Table 6 for the firms that unbunched in 2006 is similar, 40 percent of them declare taxes higher than the withheld amount. The evidence on the pattern of unbunching thus contradicts the prediction of the LPT model for about 35-40 percent of the firms.

Take together, the evidence presented above in this section suggests that a substantial proportion of firms (about 40 percent) does not behave according to the linear prospect theory model where the withholding threshold is a reference point for the taxpayer. Their tax choices are more consistent with an extended Allingham-Sandmo model where the withholding threshold is a reference point for audit. On the other hand, the evidence on the pattern of unbunching also cannot reject the linear prospect theory model for about 60 percent of the firms.

8. Conclusions

A substantial body of recent economic literature uses bunching of economic agents at policy thresholds to understand optimizing behavior. We contribute to this literature both theoretically and empirically in the context of bunching of profit tax payments by firms at the withholding threshold. Using quasi-experimental variations in withholding rates across industries and over time, we show that the bunching observed at the withholding threshold reflects primarily optimizing choices of firms rather than a lump in the distribution of profits due to lumps in the distribution of technology and risk preference. We find that the bunching around the 1 percent threshold observed in the years 2004, 2005 and 2006 vanished in 2007 for only the firms in those industries where the withholding rate changed from 1 percent to 2 percent in June 2007.

We develop alternative models of bunching of firm's tax payments that focus on three mechanisms of bunching and derive a set of testable predictions. A transaction costs model where the firms bunch because the costs of claiming refunds are higher than the amount of refund predicts that none of the firms that bunch would increase its tax payments when withholding

rate increases. The evidence based on a quasi-experimental increase in the withholding rate from 1 percent to 2 percent in 2007 in only some industries shows that the firms increased their tax payments facing a higher withholding rate. This rejects the transactions costs model.

Our analysis shows that it is not possible to generate bunching in tax payments in a prospect theoretic model with the standard S-shaped power value function of Kahneman and Tversky (1979), the withholding threshold as the reference point and loss aversion. A firm's tax payments can bunch at the withholding threshold when preferences are piece-wise linear at both sides of the reference threshold. The linear prospect theory model yields testable predictions about intertemporal persistence in bunching and the pattern of unbunching which we test with panel data. The evidence shows that a large proportion of firms that bunch in a given year unbunch in the following year which is inconsistent the prediction of intertemporal persistence in bunching in the LPT model. Among the firms that unbunch from the withholding threshold, about 35-40 percent firms declare taxes *more* than the withheld amount suggesting that the behavior of these firms is not consistent with the LPT model.

The analysis and evidence presented in this paper suggest that while the behavior of many firms conforms to the predictions of the LPT model, there is a substantial proportion of firms that behaves more according to an expected utility model of evasion where firms face higher audit risk for claiming positive refunds. The common practice of relying on a single model of evasion to understand tax choices of firms and to derive cost-benefit estimates of tax policy may not be appropriate.

References

[1] Alm, J, J. Martinez-Vazquez, and S. Wallace (2004). Taxing the Hard to Tax: Lessons from Theory and Practice. Elsevier Science.

[2] Allingham, M. G. and A. Sandmo (1972). "Income tax evasion: a theoretical analysis." *Journal of Public Economics* 1, 323-338.

[3] al-Nowaihi, Ali & Bradley, Ian & Dhami, Sanjit, (2008). "A note on the utility function under prospect theory," Economics Letters, Elsevier, vol. 99(2), pages 337-339, May.

[4] Andreoni, J., B. Erard, and J. Feinstein (1998). "Tax Compliance." *Journal of Economic Literature* 36, 818-860.

[5] Arrow, K, J (1993). "Innovation in Large and Small Firms," Journal of Entrepreneurial Finance, Pepperdine University, Graziadio School of Business and Management, vol. 2(2), pages 111-124, Spring.

[6] Asanuma, B., and T. Kikutani (1992): Risk Absorption in Japanese Subcontracting: A Microeconometric Study of the Automobile Industry, Journal of the Japanese and International Economies, 6(1), 129.

[7] Bachas, P and M. Soto (2017), "Not(ch) Your Average Tax System: Corporate Taxation Under Weak Enforcement", Working Paper, World Bank, DECRG.

[8] Becker, G. (1968). "Crime and Punishment-An Economic Approach." *Journal of Political Economy* 76, 169-217.

[9] S. Benartzi, R. Thaler (1995), Myopic Loss Aversion and the Equity Premium Puzzle, The Quarterly Journal of Economics 110, 74—92.

[10] Boadway, R. and M. Sato (2009), "Optimal Tax Design and Enforcement with an Informal Sector." *American Economic Journal: Economic Policy* 1, 1-27.

[11] Brockmeyer, Anne and Marco Hernandez (2016), "Taxation, Information, and Withholding", World Bank Policy Research Working paper No. 7600.

[12] Carrillo, P, & M. Shahe Emran, & Anita Rivadeneira, (2011). "Do Cheaters Bunch Together? Profit Taxes, Withholding Rates and Tax Evasion," Working Papers 2011-03, The George Washington University, Institute for International Economic Policy.

[13] Chetty, R, Friedman, J N., Olsen, T, & Pistaferri, L (2011). "Adjustment Costs, Firm Responses, and Micro vs. Macro Labor Supply Elasticities: Evidence from Danish Tax Records." Quarterly Journal of Economics, 126, 749804.

[14] Dhami, Sanjit, (2016). *The Foundations of Behavioral Economic Analysis,* OUP Catalogue, Oxford University Press,

[15] Dhami, Sanjit & al-Nowaihi, Ali, (2007). "Why do people pay taxes? Prospect theory versus expected utility theory," Journal of Economic Behavior & Organization, Elsevier, vol. 64(1), pages 171-192, September.

[16] Emran. M. S. and J. E. Stiglitz. (2007)."Equity and Efficiency in Tax Reform in developing Countries." Working Paper, available at SSRN.

[17] Emran, M. S. and J. E. Stiglitz (2005). "On Selective Indirect Tax Reform in Developing Countries." *Journal of Public Economics* 89, 599-623.

[18] Engel, E. M. R. A., R. D. Fischer, and A. Galetovic (1998): "Least-Present-Value-of-Revenue Auctions and Highway Franchising, NBER Working Paper No. 6689

[19] Engstrom, P, K. Nordblom, H. Ohlsson, and A. Persson (2015), "Tax Compliance and Loss Aversion", American Economic Journal: Economic Policy, November 2015.

[20] Goode, R. (1984), Government Finance in Developing Countries. Brookings Institution.

[21] Gordon, R. ed. (2010). Taxation in Developing Countries: Six Case Studies and Policy Implications. Columbia University Press.

[22] Gordon , R. and W. Li, (2009). "Tax Structure in Developing Countries: Many Puzzles and a Possible Explanation." *Journal of Public Economics*, 93, 855-866.

[23] Greenwald, Bruce C and Stiglitz, Joseph, (1990), "Asymmetric Information and the New Theory of the Firm: Financial Constraints and Risk Behavior," American Economic Review, 80, issue 2, p. 160-65.

[24] Greenwald, B, and J. Stiglitz (2014), "Creating a Learning Society: A New Approach to Growth, Development, and Social Progress," Columbia University.

[25] Herranz, N, S. Krasa, and A. Villamil (2015), "Entrepreneurs, Risk Aversion and Dynamic Firms", Journal of Political Economy, October 2015.

[26] Kahneman, Daniel & Tversky, Amos, (1979). "Prospect Theory: An Analysis of Decision under Risk," Econometrica, Econometric Society, vol. 47(2), pages 263-291, March.

[27] Kapoor, S (2017), "Are Firms Loss Averse? Evidence from the Restaurant Industry", Working paper, Erasmus University. Accessed on June 25, 2018.

[28] Keen, M. and J. Mintz (2004). "The Optimal Threshold for a Value-Added Tax." *Journal of Public Economics* 88, 559-576.

[29] Khwaja, A, & Atif Mian, (2008). "Tracing the Impact of Bank Liquidity Shocks: Evidence from an Emerging Market," American Economic Review, American Economic Association, vol. 98(4), pages 1413-1442, September.

[30] Kleven, H. J., M. B. Knudsen, C. T. Kreiner, S. Pedersen and E. Saez (2011). "Unwilling or Unable to Cheat? Evidence From a Randomized Tax Audit Experiment in Denmark." *Econometrica*, Econometric Society, vol. 79(3), pages 651-692, May.

[31] Kleven, H, & Claus Thustrup Kreiner, & Emmanuel Saez, 2016. "Why Can Modern Governments Tax So Much? An Agency Model of Firms as Fiscal Intermediaries," Economica, London School of Economics and Political Science, vol. 83(330), pages 219-246.

[32] Kleven, H, & Waseem, M (2013). "Using Notches to Uncover Optimization Frictions and Structural Elasticities: Theory and Evidence from Pakistan." Quarterly Journal of Economics, 128, 669723.

[33] Kleven, H (2016), "Bunching", Annual Review of Economics, 8, 435-464, 2016.

[34] Lind, Jo and Mehlum, Halvor, (2010), With or Without U? The Appropriate Test for a U-Shaped Relationship, Oxford Bulletin of Economics and Statistics, 72, issue 1, p. 109-118.

[35] Neilson, W (2002), "Comparative Risk Sensitivity with Reference-Dependent Preferences," Journal of Risk and Uncertainty, 24:2; 131142, 2002.

[36] Onji, K. (2009). "The response of firms to eligibility thresholds: Evidence from the Japanese value-added tax." *Journal of Public Economics* 93, 766-775.

[37] Rabin, M (2000). "Risk Aversion and Expected-Utility Theory: A Calibration Theorem," Econometrica, Econometric Society, vol. 68(5), pages 1281-1292, September.

[38] Rees-Jones, A (2017), "Quantifying Loss-Averse Tax Manipulation", Review of Economic Studies, 2017.

[39] Reinganum, J. and Wilde, L. (1985). "Income Tax Compliance in a Principal-Agent Framework." *Journal of Public Economics* 26, 1-18.

[40] Reinganum, J. and Wilde, L. (1986), "Equilibrium Verification and Reporting Policies in a Model of Tax Compliance." *International Economic Review* 27, 739-760.

[41] Sabine, B. (1966). A History of Income Tax. George Allen, London.

[42] Saez, E. (2010). "Do Taxpayers Bunch at Kink Points?" *American Economic Journal: Economic Policy* 2, 180-212.

[43] Sandmo, A. (2005). "The Theory of Tax Evasion: A Retrospective View." *National Tax Journal* 58, 643-63.

[44] Slemrod, J (2017), "Tax Compliance and Enforcement: An Overview of New Research and Its Policy Implications." In Alan Auerbach and Kent Smetters, eds., The Economics of Tax Policy, Oxford University Press, 2017, pp. 81-102.

[45] Selmrod, J (2010), Buenas Notches: Lines and Notches in Tax System Design, Working paper, University of Michigan.

[46] Slemrod, J. and S. Yitzhaki (2002). "Tax avoidance, evasion and administration," in A.J. Auerbach and M. Feldstein (eds.), Handbook of Public Economics, Vol. 3, Elsevier: Amsterdam.

[47] Soos, P. (1990). "Self-employed Evasion and Tax Withholding: A Comparative Study and Analysis of the Issues." U.C. Davis Law Review.

[48] Stiglitz, J. (2010), Development Oriented Tax Policy, in Gordon, R. ed. (2010).

[49] Yaniv G., (1999). "Tax compliance and advance tax payments: A prospect theory analysis." National Tax Journal 52, 753-64.

[50] Yaniv, G. (1988). "Withholding and Non-Withheld Tax Evasions." *Journal of Public Economics* 35, 183-204.

[51] Yitzhaki, S. (1974). "A Note on Income Tax Evasion: A Theoretical Analysis." *Journal of Public Economics* 3, 201-202.

Table 1: Profit-Tax Withholding Rates (as a Fraction of Sales)

Industry	Jan. 04 to June 07	July 07 to Dec. 07
Construction	0.01	0.02
Manufacturing	0.01	0.02
Wholesale and Retail	0.01	0.02
Transportation	0.01	0.01

Notes: The withholding rates are reported for the period 2004-2007 that correspond to our panel data. The rates generally apply to the typical firm in each of the four industries above.

Table 2: Characteristics of the Firms (Balanced Panel: 2006-2007)

Variable Name	Description	2006		2007	
		Mean	St. Dev.	Mean	St. Dev.
Sales	Total sales ($ thousands) reported by firms in tax return.	2,674.4	26,813.9	3,051.7	28,520.6
Tax	Profit-tax ($ thousands) declared by firms in tax return.	25.7	331.3	31.9	366.9
Assets	Total assets ($ thousands) reported by firms in tax return.	1,556.5	12,304.7	1,737.7	13,982.9
Tax / Sales	Reported profit-tax as a share of total sales.	0.009	0.019	0.011	0.020
Manufacturing	Equals 1 if Manufacturing firm.	0.211	0.408	0.211	0.408
Construction	Equals 1 if Construction firm.	0.093	0.290	0.093	0.290
Wholesale / Retail	Equals 1 if Wholesale or Retail Trade firm.	0.570	0.495	0.570	0.495
Transportation	Equals 1 if firm Transportation firm.	0.126	0.332	0.126	0.332

Table 3: Do Firms Bunch Near the Withholding Rate?

A) Saez (2010) Test for Bunching During 2004 - 2006 (p-values)

	B_a 1(|tax/sales - 0.01|< .0008)	B_b 1(|tax/sales - 0.01|< .0009)	B_c 1(|tax/sales - 0.01|< .0010)
Contruction	2.6E-18	1.3E-20	1.5E-24
Manufacturing	5.4E-04	7.2E-05	4.2E-06
Retail	3.6E-10	5.3E-13	5.1E-14
Transportation	7.1E-07	1.7E-05	3.9E-06

B) Saez (2010) Test for Bunching During 2006 (p-values)

	B_a 1(|tax/sales - 0.01|< .0008)	B_b 1(|tax/sales - 0.01|< .0009)	B_c 1(|tax/sales - 0.01|< .0010)
Contruction	3.0E-05	8.2E-08	5.0E-12
Manufacturing	0.0099	0.0022	0.0002
Retail	1.1E-06	3.5E-07	1.9E-07
Transportation	2.5E-05	4.8E-04	0.0002

C) Saez (2010) Test for Bunching During 2007 (p-values)

	B_a 1(|tax/sales - 0.01|< .0008)	B_b 1(|tax/sales - 0.01|< .0009)	B_c 1(|tax/sales - 0.01|< .0010)
Construction	0.8618	0.4571	0.5342
Manufacturing	0.7895	0.7618	0.2948
Retail	0.2105	0.7697	0.4771
Transportation	0.0598	0.0137	0.0035

Notes: This Table displays p-values of a statistical test under the null hypothesis of "no-bunching" near the 2006 withholding rate of 0.01. A firm "bunches" if its reported tax/sales ratio lies within a small distance (ρ) from the withholding rate. For robustness, we choose three alternative values for ρ (0.08%, 0.09% and 0.10%) that are displayed in each of the columns of the Table. The variable B_a equals one if firm i's tax/sales ratio in 2006 is between 0.92% and 1.08%. Similarly, the variables B_b and B_c equal one if the 2006 tax/sales ratio falls within the intervals [0.91%, 1.099%] and [0.90%, 1.10%], respectively. For each of the three bands defined above, two symmetric surrounding intervals are computed: [1%-2ρ, 1%-ρ] and [1%+ρ, 1%+2ρ]. P-values are reported above.

Table 4: Changes in Reported Taxes and Withholding Rates

Dependent Variable is the Change in the Firm's Profit Tax / Sales Ratio Between 2007 and 2006

	Change in δ Yes		Change in δ No
B_a: 1 (| tax/sales - δ| < .0008)	0.0010	**	0.0000
	(0.0004)		(0.0015)
B_b: 1 (| tax/sales - δ| < .0009)	0.0010	**	0.0004
	(0.0004)		(0.0014)
B_c: 1 (| tax/sales - δ| < .0010)	0.0011	***	0.0019
	(0.0004)		(0.0017)
Number of Observations	**12,082**		**1,748**

Notes: (1) The estimates show the coefficient of the dummy variable for bunching from a DID design with the firms that did not bunch as the comparison group. δ denotes the withholding rate. The first column reports the estimates for those industries where the withholding rate increased from 1 to 2 percent in 2007. The second column shows estimates for firms with unchanged withholding rate. (2) Regressions include change in assets between 2006 and 2007 as a control. (3) *** and ** denote statistical significance at the 1 and 5 percent level, respectively; estimates without any asterisk are not statistically different from zero.

Table 5A: Bunching and Firm Size: Large Firm Effect

(The dependent variable is a dummy that takes on the value 1 when

a firm bunches at the withholding threshold in 2006.)

	Ba		Bb		Bc	
1 (Assets (t-1) > p75)	-0.015	***	-0.016	***	-0.017	***
	(0.0036)		(0.0038)		(0.0039)	
1 (Assets (t-1) > p90)	-0.025	***	-0.030	***	-0.030	***
	(0.005)		(0.005)		(0.005)	
1 (Assets (t-1) > p95)	-0.034	***	-0.040	***	-0.0409	***
	(0.006)		(0.006)		(0.0065)	
1 (Assets (t-1) > p99)	-0.0481	***	-0.0569	***	-0.0583	***
	(0.0112)		(0.0112)		(0.0121)	
Number of Observations	28,828		28,828		28,828	

Notes: This Table shows results of univariate regressions where the dependent variable (in each column) is independently regressed on the covariates in each row. Assets are measured as the average of 2004 and 2005. The independent variable in each regression is an indicator that takes the value of one if a firm's assets are above a specific percentile. For example, 1 (Assets (t-1) > p75) denotes the firms that belong to the right tail above 75th percentile of the distribution. *** denotes statistical significance at the 1 percent level.

Table 5B: Bunching and Firm Size: Inverted-U?

(The dependent variable is a binary indicator of bunching that takes 1 when a firm bunches near the withholding threshold in 2006.)

|	B_a		B_b		B_c	
Firm Size Rank (t-1)	**0.125**	***	**0.127**	***	**0.147**	***
	(0.0218)		**(0.0230)**		**(0.0238)**	
(Firm Size Rank)² (t-1)	**-0.140**	***	**-0.146**	***	**-0.164**	***
	(0.0211)		**(0.0221)**		**(0.0230)**	
Slope at the Lower Bound (1st Percentile)	**0.122**		**0.124**		**0.144**	
	(0.0214)	***	**(0.0225)**	***	**(0.0233)**	***
Slope at the Upper Bound (99th Percentile)	**-0.152**		**-0.162**		**-0.177**	
	(0.0214)		**(0.0223)**		**(0.0232)**	
Sasabuchi t test	**5.70**		**5.52**		**6.61**	
P-Value	**6.09E-09**		**1.68E-08**		**3.72E-10**	
Extremum Point	**0.446**		**0.435**		**0.449**	
95 Percent Conf. Interval (Fieller Method)	**[0.397,0.485]**		**[0.383,0.475]**		**[0.404,0.485]**	
Number of observations	28,828		28,828		28,828	

Notes: (1) Firm size is measured in percentile ranks in asset distribution. The assets are measured as the average of 2004 and 2005. (2) *** denotes statistical significance at the 1 percent level. (3) The null hypothesis for Sasabuchi t test is monotonic relation or U-shape, and the alternative is inverted-U shape.

Table 6: Persistence in Bunching and the Pattern of Unbunching

|	B_a	B_b	B_c
Persistent Bunchers (percent)	**3.2**	**3.6**	**4.5**
Firms that unbunch in 2005 (percent) (Among those who bunched in 2004)	**70.39**	**68.61**	**66.47**
Percent with taxes more than withheld amount in 2005 (among those who unbunch in 2005)	**35.48**	**35.41**	**35.55**
Firms that unbunch in 2006 (percent) (Among those who bunched in 2005)	**70.1**	**68.42**	**65.87**
Percent with taxes more than withheld amount in 2006 (Among those who unbunched in 2006)	**39.17**	**38.7**	**38.21**

Notes: The definitions of bunching in three columns are same as in Table 3.

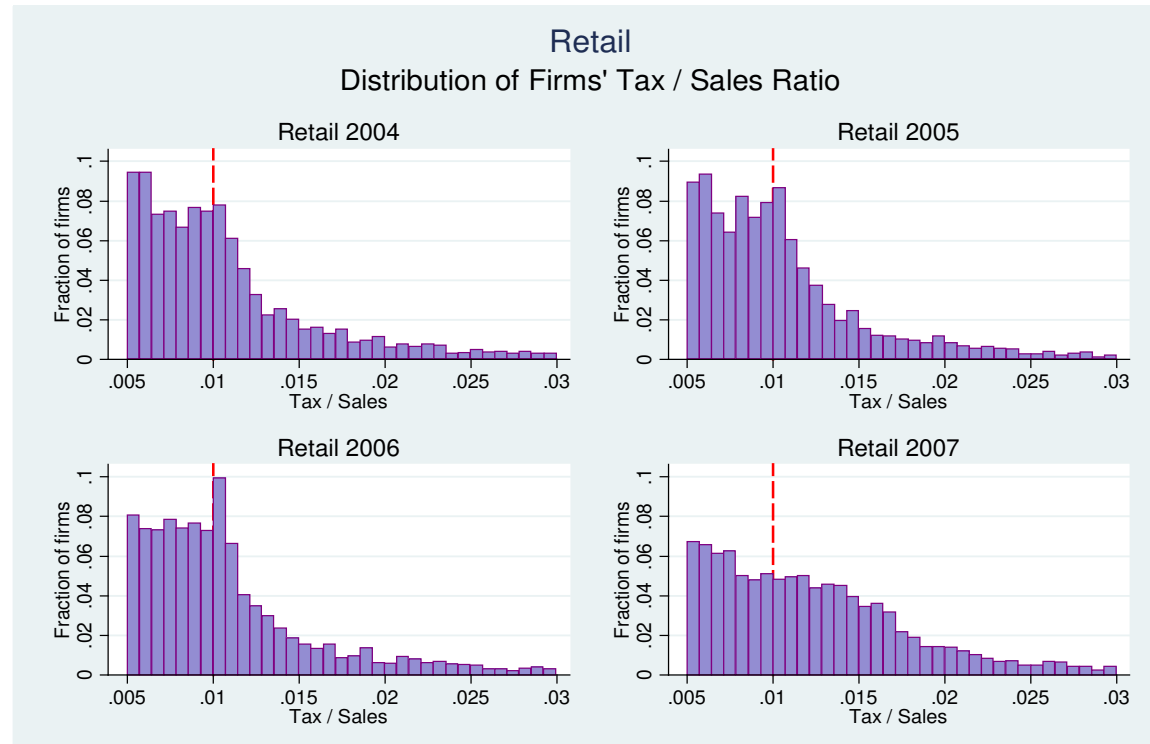## Figure 1: Profit-Taxes Reported by Construction Firms
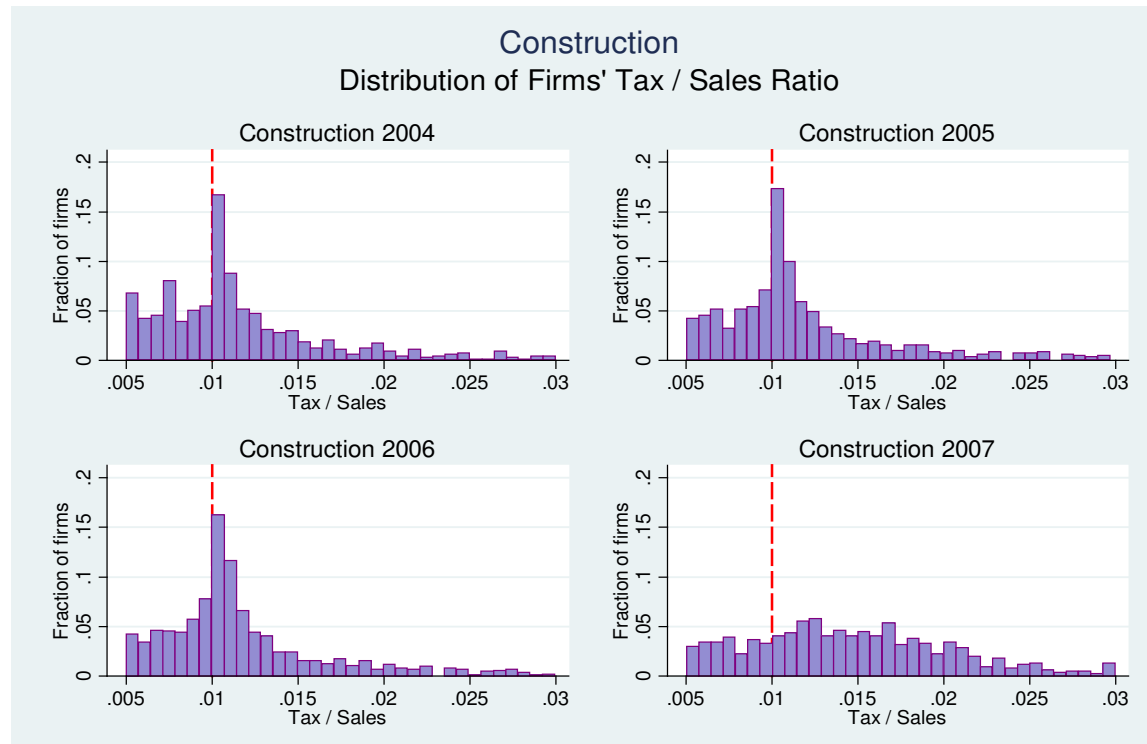
Note: Figures show histograms for the tax / sales ratio of Ecuadorian firms in the Construction sector by year. The vertical dashed line shows the witholoding rate (1%) that applied to firm's sales in 2004, 2005 and 2006. On average, there are 1,449 Construction firms per year in our sample. Part of the distribution has been left out of the graph.

Figure 2: Profit-Taxes Reported by Manufacturing Firms
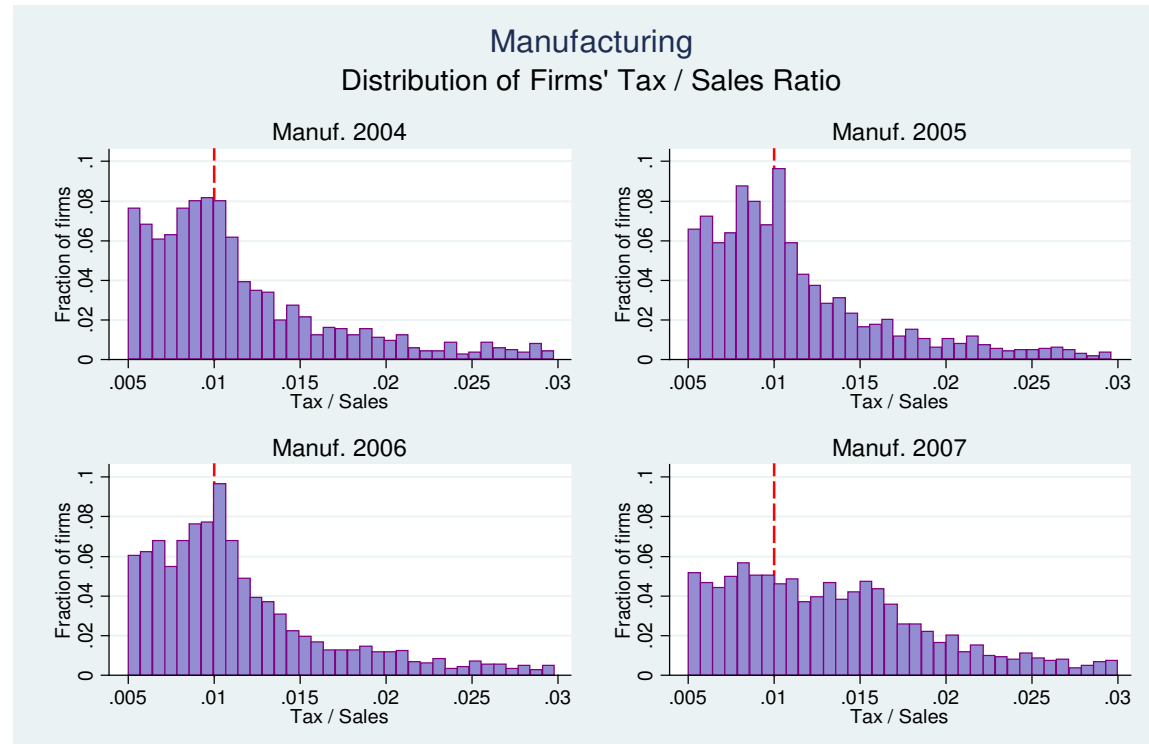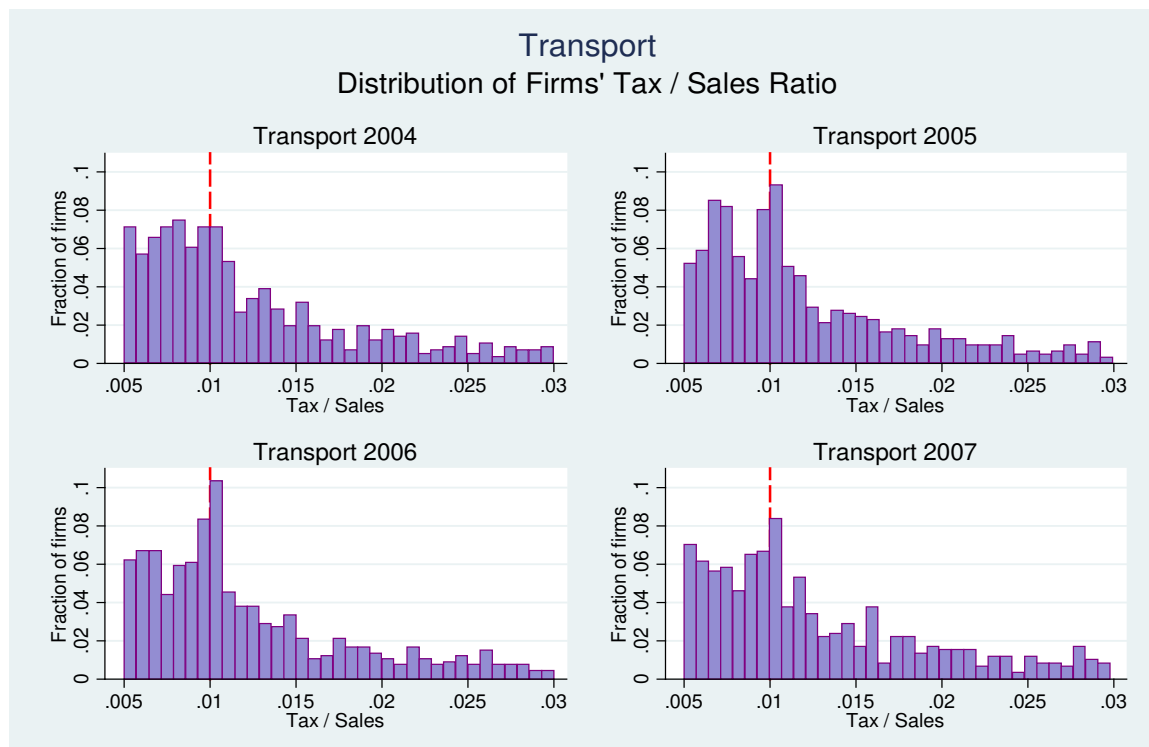

Note: Figures show histograms for the tax / sales ratio of Ecuadorian firms in the Manufacturing sector by year. The vertical dashed line shows the witholoding rate (1%) that applied to firm's sales in 2004, 2005 and 2006. On average, there are 3,192 Manufacturing firms per year in our sample. Part of the distribution has been left out of the graph.

**Figure 3: Profit-Taxes Reported by Wholesale / Retail Firms**

Note: Figures show histograms for the tax / sales ratio of Ecuadorian firms in the Retail sector by year. The vertical dashed line shows the witholoding rate (1%) that applied to firm's sales in 2004, 2005 and 2006. On average, there are 8,705 Retail firms per year in our sample. Part of the distribution has been left out of the graph.

Figure 4: Profit-Taxes Reported by Transport Firms

Note: Figures show histograms for the tax / sales ratio of Ecuadorian firms in the Retail sector by year. The vertical dashed line shows the witholoding rate (1%) that applied to firm's sales in 2004, 2005 and 2006. On average, there are 2,193 Retail firms per year in our sample. Part of the distribution has been left out of the graph.

Figure 5: Share of Firms that Bunch by Asset Decile
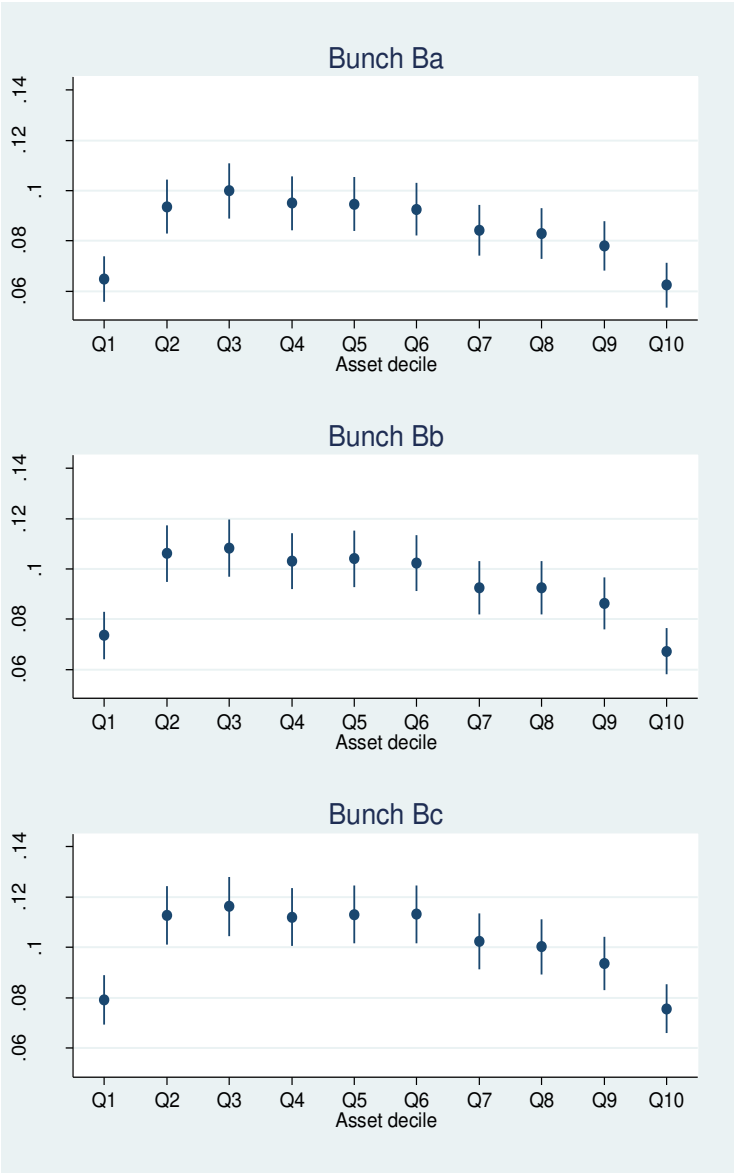
Online Appendix: Not for Publication

Proof of Proposition 1

(1.a) Note first that in the model above it is never optimal for a firm to declare that $\pi_i^* > \Pi_i$, because while a taxpayer pays a penalty for underreporting tax liability, they get back \$1 in case of audit when \$1 is overreported. So we need to show that there exists $\hat{P} \in (0,1)$ such that $\forall P \geq \hat{P}$, it is not optimal to declare that $\pi_i^* < \Pi_i$. Observe that $\pi_i \geq 0$, as a firm when making loss cannot ask for negative tax payments. For $\pi_i^* = \Pi_i$, the following condition needs to be satisfied:

$$\frac{dW_i}{d\pi_i}\Big|_{\pi_i=\Pi_i} \geq 0 \implies \left[\frac{U_i'\left(Z_i^b\right)}{U_i'\left(Z_i^g\right)}\right]_{\pi_i=\Pi_i} \geq \frac{(1-P)}{P\theta}$$

$$\iff 1 \geq \frac{(1-P)}{P\theta}, \text{ because } \pi_i = \Pi_i \implies Z_i^b = Z_i^g, \ \forall \ i$$

$$\iff P \geq (1+\theta)^{-1} \equiv \hat{P}$$

(1.b) For $\pi_i^* = 0$, the following condition needs to be satisfied:

$$\frac{dW_i}{d\pi_i}\Big|_{\pi_i=0} \leq 0 \implies \left[\frac{U_i'\left(Z_i^b\right)}{U_i'\left(Z_i^g\right)}\right]_{\pi_i=0} \leq \frac{(1-P)}{P\theta}$$

$$\implies P \leq (1+\theta\eta_i)^{-1} \equiv \tilde{P}_i, \quad \text{where } \eta_i \equiv \left[\frac{U_i'\left(Z_i^b\right)}{U_i'\left(Z_i^g\right)}\right]_{\pi_i=0}$$

To see that $\hat{P} > \tilde{P}_i$, note that

$$\eta_i \equiv \left[\frac{U_i'\left(Z_i^b\right)}{U_i'\left(Z_i^g\right)}\right]_{\pi_i=0} = \left[\frac{U_i'\left(\Pi_i\left\{1-\tau\left(1+\theta\right)\right\}\right)}{U_i'\left(\Pi_i\right)}\right] > 1$$

The last inequality above follows from the facts that $U_i(.)$ is concave and $\{1 - \tau(1+\theta)\} < 1$.

(1.c) For a risk neutral firm, $\pi_i^* = 0$ is the corner solution when $P < \hat{P}$. Now consider a risk averse firm, when $P \in \left(\tilde{P}_i, \ \hat{P}\right)$. The first order condition in equation (3) of the main text is satisfied in this case with $\pi_i^* \in (0, \Pi_i)$. This last statement follows immediately from propositions (1.a) and (1.b). The last part of (1.c) that the optimal tax declared is higher for a more risk averse firm follows from the first order condition, and noting that if firm i_1 is more risk averse than firm i_0, then

$$U_{i_0}'(Z) < U_{i_1}'(Z).$$

1

Proof of Proposition 2

By definition

$$\mu_i \equiv \left[\frac{U_i'\left(Z_i^g\right)}{U_i'\left(Z_i^b\right)}\right]_{\pi_i \tau = \delta Y_i}$$

In other words, μ_i is the ratio of marginal utility in bad to good states evaluated at the reported cost that makes the declared taxes exactly equal to the withheld amount. The lower threshold $\tilde{\mu}_i$ is defined by the following condition:

$$\left.\frac{dW_i}{dC_i^d}\right|_{\pi_i = \frac{\delta Y_i}{\tau}} \geq 0 \implies \mu_i \geq \tilde{\mu}_i \equiv \frac{P_2 \theta}{(1 - P_2)}$$

Thus $\tilde{\mu}_i$ is the critical value that determines whether a firm finds it optimal to report costs high enough to make the refund positive when facing probability of audit P_2. First consider the case when $P_1 > \hat{P}$. In this case for any firm with $\tilde{\Psi}_i > \delta$ and $\mu_i \geq \tilde{\mu}_i$ bunch at the withholding threshold. The proof follows from an application of the results in proposition (1) and noting that given the discontinuity in the audit function at the threshold, it is not optimal for a firm with $i \ni \mu_i \geq \tilde{\mu}_i$ to claim a tax liability lower than the threshold, because it is optimal to declare $\pi_i^* = \Pi_i$ facing probability of audit $P_1 > \hat{P}$.

When $P_1 < \hat{P}$, proposition (1.c) implies that any risk neutral firm declares $\pi_i^* = 0$, and among the risk averse firms there is a subset that finds it optimal to evade taxes and declare $\pi_i^* \in (0, \Pi_i)$. This implies that some risk averse firms that bunched at the withholding threshold under the case $P_1 > \hat{P}$ would now declare tax liability below the threshold facing $P_1 < \hat{P}$. To identify the set of firms that bunch at the threshold we thus need to define a upper bound such that the following holds:

$$\left.\frac{dW_i}{dC_i^d}\right|_{\pi_i = \frac{\delta Y_i}{\tau}} \leq 0 \implies \mu_i \leq \breve{\mu}_i \equiv \frac{P_1 \theta}{(1 - P_1)}$$

Thus the set of firms that bunch at the threshold under the assumption that $P_1 < \hat{P}$ is defined as $i \in (\tilde{\mu}_i, \breve{\mu}_i)$.