NBER WORKING PAPER SERIES

TAX AUDITS AS SCARECROWS:
EVIDENCE FROM A LARGE-SCALE FIELD EXPERIMENT

Marcelo L. Bérgolo
Rodrigo Ceni
Guillermo Cruces
Matias Giaccobasso
Ricardo Perez-Truglia

Working Paper 23631
http://www.nber.org/papers/w23631

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
July 2017, Revised October 2018

We thank the Tax Registry of Uruguay for their collaboration. We thank Gustavo Gonzalez for his support, without which this research would not have been possible. We thank Joel Slemrod for his valuable feedback, as well as that of seminar participants at the University of Michigan, Universidad Di Tella, Universidad de la Republica, the 2017 NBER Public Economics Fall Meeting, the 2017 RIDGE Public Economics Conference, the 2017 Zurich Center for Economic Development Conference, the 2017 Advances with Field Experiments Conference, the 2018 PacDev Conference, the 2018 AEA Annual Meetings and the 2018 IIPF annual Congress. This project benefited from funding by CEF, CEDLAS-UNLP and IDRC. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Tax Audits as Scarecrows: Evidence from a Large-Scale Field Experiment
Marcelo L. Bérgolo, Rodrigo Ceni, Guillermo Cruces, Matias Giaccobasso, and Ricardo Perez-Truglia
NBER Working Paper No. 23631
July 2017, Revised October 2018
JEL No. C93,H26,K42

ABSTRACT

The canonical model of Allingham and Sandmo (1972) predicts that firms evade taxes by optimally trading off between the costs and the benefits of evasion. However, there is still no consensus about whether real-world firms react to audits in this way. We conducted a large scale field experiment in collaboration with Uruguay's tax authority to shed light on these issues. We sent letters to 20,440 small-and medium-sized firms that collectively pay over 200 million dollars in taxes per year. Our letters provided exogenous yet nondeceptive signals about key inputs for their evasion decisions such as audit probabilities and penalty rates. We measure the effect of these signals on their subsequent perceptions about the auditing process, based on survey data, as well as on the actual taxes paid, according to administrative data. We find that firms increase their tax compliance in response to information about audits. However, we do not find these effects to be consistent with Allingham and Sandmo (1972). Our favorite interpretation of our findings is based on the model of risk-as-feelings: audits may deter tax evasion in the same way that scarecrows frighten off birds.

Marcelo L. Bérgolo
Instituto de Economia (IECON)
Universidad de La Republica
1375 Joaquin Requena
Montevideo, Uruguay
mbergolo@iecon.ccee.edu.uy

Rodrigo Ceni
Instituto de Economia (IECON)
Universidad de La Republica
1375 Joaquin Requena
Montevideo, Uruguay
rceni@iecon.ccee.edu.uy

Guillermo Cruces
CEDLAS
Univesidad Nacional de La Plata
Calle 6 entre 47 y 48
La Plata, Argentina
gcruces@cedlas.org

Matias Giaccobasso
Anderson School of Management - UCLA
110 Westwood Plaza, C 3.10
Los Angeles, CA 90077
matias.giaccobasso.phd@anderson.ucla.edu

Ricardo Perez-Truglia
Anderson School of Management
University of California, Los Angeles
110 Westwood Plaza
Los Angeles, CA 90095
and NBER
ricardo.truglia@anderson.ucla.edu

1 Introduction

Tax audits have been standard tools of most tax administrations throughout history. While in all likelihood audits are useful to foster tax compliance, there is no consensus yet on exactly how firms react to them. The standard economic model posits that they affect firms' behavior through the expected value of evasion fines, but other factors might be at play. We present evidence from a large scale field experiment designed to tackle these questions.

Audits increase tax revenues mechanically, since firms caught evading must pay taxes on the hidden income, as well as evasion penalties. However, audits demand resources, and in all likelihood auditing only to obtain these direct revenues is not cost efficient. In fact, tax administrations concentrate most of their auditing and supervision efforts on a small number of large firms, leaving smaller firms relatively alone. The fact that tax administrations conduct audits at all on the latter group of small, low revenue firms, is probably due to their deterrence effect: firms are more likely to report their incomes to the authorities because of the threat of being audited in the future (and its expected costs). The canonical economic model of tax evasion, that of Allingham and Sandmo (1972) (hereon after, *A-S*), posits that firms respond to audits and their deterrent effect in an optimal way: they choose the optimal amount of income to hide from the tax authority so that the marginal benefits (i.e., the lower tax burden) equate the marginal costs (i.e., the penalties to be paid if caught).[1]

Whether firms in the real-world react to audits in the calculated manner predicted by *A-S* is still a source of debate Alm, Jackson, and McKee (1992); Dhami and al Nowaihi (2007); Luttmer and Singhal (2014); Slemrod (2018). This is an important question for at least two reasons. First, understanding the true nature of firm behavior in this aspect can also shed light into the broader nature of decision making and profit maximization at the firm level. Second, the answer to this question has a number of implications, for instance for the design of audit strategies and to explain why some countries exhibit higher tax compliance than others.

We collaborated with the Internal Revenue Service (IRS) from Uruguay to conduct a field experiment based on information about audits with a sample of 20,440 small- and medium-size firms. These firms' largest tax liability is the Value Added Tax (VAT). Unlike other sources of taxable income, such as wage income, the VAT has limited third party reporting,[2] and thus tax authorities can only detect VAT evasion in small or medium firms through audits.

[1]This is an application of the more general Becker (1968) model of the economics of crime, in which selfish and risk-averse individuals choose whether to engage in criminal activities according to the same trade off.

[2]For instance, wage income is hard to conceal in modern economies because employers are usually required to report their employees' earnings to the tax authority. As a result, tax evasion can be detected automatically and thus deterred without the need to conduct audits (Kleven, Knudsen, Kreiner, Pedersen, and Saez (2011)).

In other words, audits are the primary mechanism through which tax administrations can detect VAT evasion in small and medium enterprises (Gomez-Sabaini and Jimenez (2012); Bergman and Nevarez (2006)), and thus in this context such firms should be aware and pay attention to audits and their consequences.[3]

The IRS mailed different types of letters with information about audits to the owners of each of these firms.[4] We randomly assigned the information contained in each of these letters to test specific hypothesis about the role of audits in tax compliance. We measure the subsequent effects of the information contained in the letters on VAT and other taxes paid by these firms obtained from IRS administrative records. Additionally, we conducted a post-mailing survey to capture the effect of this information on these firms' subsequent perceptions about audits – the perceived probability of being audited and the perceived penalty rate.

The first part of the experimental design is set up to confirm in this context a standard finding in the literature for other countries: communicating information about enforcement increases tax compliance (Slemrod, Blumenthal, and Christian (2001)). Firms were randomized into different letter types. The *baseline* letter type included brief generic information about taxes that the IRS often includes in its communications with firms. The *audit-statistics* letter type was identical to the *baseline* letter, with the exception of an additional message with information about the probability of being audited and about the penalty rate, based on tax administration statistics. Similarly, the the *audit-endogeneity* letter was identical to the *baseline* letter, except for an additional message conveying another aspect of the auditing process: evading taxes increases the probability of being audited. The last letter type was designed to provide a benchmark for a message that might increase tax compliance but that is not related to audits: the *public-goods* letter was identical to the *baseline* letter, except for an additional message listing all the public goods that could be provided if firms reduced evasion by 10%.

The relevant hypothesis is that the post-treatment tax payments of firms assigned to the *baseline* letters will be lower than those of recipients of *audit-statistics* letters. Additionally, we can compare the effects of this *audit-statistics* message to the effects of the other messages, some related to audits (*audit-endogeneity*) and some unrelated (*public-goods*). The results are broadly consistent with the previous evidence: the information about audits (especially the *audit-statistics* message) significantly increased tax compliance.

[3]While the VAT requires a paper trail, it is subject to significant limitations. For once, the paper trail breaks down when reaching the consumer. Moreover, firms can collude with each other, or with customers, to avoid or tamper with the paper trail. This type of collusion and the resulting is relatively frequent in middle income countries and particularly in Latin America – see Pomeranz (2015) for Chile and Naritomi (2016) for Brazil.

[4]Throughout the paper, for simplicity, we refer to firms' perceptions and behavior as a shorthand for firms' owners or managers perceptions and behavior.

The second part of the experimental design tries to shed light on the mechanism underlying the effects of the audit messages. More precisely, we test the hypothesis that firms react to the information about audits following the cost–benefit calculation posited by *A-S*. We test this hypothesis in three different ways.

First, we show that the information about audits in the letters actually changed firms' beliefs. This result is based on a purportedly-designed survey of firms in our experimental sample which elicited perceptions about the probability of being audited and about the penalty rate. We conducted this survey nine months after the experiment. Since the *audit-statistics* letters increased average compliance, to be consistent with *A-S*, the information in the letters should have increased the perceived probability of being audited and/or increased the perceived penalty rate. However, our evidence shows that the letters actually reduced the perceived audit probability.

The second test exploits the fact that *A-S* predicts that different firms' reaction to the information should be a function of their prior beliefs. For example, if a firm's prior belief about the probability of an audit was lower than that provided in the letter, the firm should increase its tax compliance – the new information indicates that the firm is more likely to be caught than what its prior implied. The opposite is true for firms with prior beliefs about probabilities of audits that are higher than the probability shown in the letter: these firms should decrease their tax compliance, because the new information implies that they are less likely to be caught evading than what their priors implied. We test this hypothesis by studying the heterogeneity of the experiment's effects among groups of firms with different priors. We construct a proxy for prior beliefs about audits based on random variation in pre-treatment exposure to audits. Firms that were audited in the past should differ in their beliefs about audits' probabilities and fines from those with no previous audit experience, at least if they follow some form of Bayesian learning. While firms audited in the past report higher perceived probabilities of audits, the effect of the letters on tax compliance is not higher for this group.

The third test is based on *A-S'* direct predictions about the effect of audit probabilities and fines on evasion: the effects on tax compliance should be increasing in the values of these parameters conveyed in our letters. We generated exogenous non-deceptive variation in information about audit probabilities and penalty rates in a series of *audit-statistics* sub-treatments. To do so, we computed the average probabilities and penalty rates using a series of random samples of 50 firms. This limited sample size introduced substantial sampling variation in average probabilities and fines, which we then included in the letters. For instance, a firm may have received a signal of audit probabilities of either 8% or 15%, depending on which underlying sample of similar firms was drawn for that particular letter. This random variation allows us to test whether the effect of the message on tax compliance is a function of the signal conveyed about the audit probability and the penalty rate.

Moreover, as a complement to the *audit-statistics* sub-treatments, we designed a treatment arm, the *audit-threat* letter, which created exogenous variation in expected audit probabilities in a more direct way. This letter type is based on a separate sample of firms pre-selected by the IRS for auditing. We randomly divided this set of firms into two groups, one that would be audited with a 25% probability and the other that would be audited with a 50% probability, and informed them (by means of similar letters) which audit probability they had been assigned. We test the hypothesis that firms randomly assigned to the 50% probability have higher tax compliance than those assigned to the 25% probability group (and message).

The first part of the results show that sending messages related to audits increases tax compliance. Adding the *audit-statistics* message to the *baseline* letter increases tax compliance by about 6.3%. This effect is economically large: using the estimated average evasion rate for VAT in Uruguay of 26% (Gomez-Sabaini and Jimenez (2012)), the 6.3% effect amounts to a reduction in the evasion rate of 24%. This effect is highly statistically significant and robust to a number of checks, such as alternative specifications and event-study falsification tests.

The other message related to audits, *audit-endogeneity*, also significantly increased tax compliance by about 7.4% – this effect is statistically indistinguishable from the 6.3% effect of the *audit-statistics* message. This effect is also robust to a number of checks. In comparison, the message unrelated to audits, the *public-goods* treatment, had a less clear effect on tax compliance: in the main specification, its effect is smaller (4.3%) and statistically insignificant, and is not robust across specifications.

The second part of the results provide evidence on whether the effects of the *audit-statistics* message is consistent with the *A-S* model. The three tests provide robust evidence against against it. The first test, based on survey data, indicates that the *audit-statistics* message actually reduced the perceived probability of being audited. Thus, if firms reacted to the *audit-statistics* letter as predicted by *A-S*, they should have reduced, rather than increased, their tax compliance.

The second test relies on our proxy for prior beliefs about the probability of being audited, derived from each firm's audit history. The survey data validates this proxy: firms recently exposed to audits report substantially higher perceived probabilities of being audited in the future. We then show that the effects of the *audit-statistics* message on tax compliance are not higher for individuals with higher prior beliefs – if anything, there is a statistically insignificant difference in the opposite direction. This directly contradicts the *A-S* model prediction.

The results from the third test, based on the sub-treatments with variation in signals about audit probabilities and fines, is also inconsistent with the core *A-S* prediction. Firms that by chance received higher signals of the audit probability or penalty rates in their *audit-statistics* letters do not pay a significantly higher amount of taxes. Indeed, the average elasticities of tax compliance with respect to the audit probability and to the penalty rate are close to zero, precisely estimated and statistically different from those predicted by various calibrations of the *A-S* model. The results are similar for firms assigned to the *audit-threat* letter, which provides an alternative source of variation to the perceived probability of being audited.

Our findings suggest that small and medium firms may comply with taxes because of the threat of being audited, but not necessarily in an optimal manner as in *A-S*. Our research design's main objective is to test this model's null hypotheses, but we discuss also discuss whether the results are consistent with alternative models. We argue that two important alternatives in the literature, salience (Chetty, Looney, and Kroft (2009)) and prospect theory (Kahneman and Tversky (1979)), are also inconsistent with our results. In contrast, the model of risk-as-feelings (Loewenstein, Weber, Hsee, and Welch (2001)) is compatible with our findings. In this setting, small and medium firms may react to messages about audits because these messages trigger an irrational feeling of fear, even they do not affect the beliefs about audit probabilities and fines. In other words, tax audits may deter taxpayers from evading taxes in the same way that scarecrows deter birds from eating crops.

Our study is related to various strands of literature. First, it belongs to a recent but growing literature that uses field experiments to study the decision of individuals to pay taxes. In a seminal contribution, Slemrod et al. (2001) showed that, for a sample of U.S. self-employed individuals, those who were randomly assigned to receive a letter from the Internal Revenue Services with an enforcement message reported higher income in their tax returns. Similar messages about tax enforcement have been shown to have positive effects on tax compliance in other contexts (for reviews of the recent literature, see Slemrod (2018) and Pomeranz and Vila-Belda (2018). Our reading of this literature is that taxpayers incorporate information about tax enforcement tools and, in line with *A-S*, reduce their evasion to re-optimize their behavior. Models of salience provide an alternative interpretation: taxpayers may behave as if the enforcement tools did not exist until these tools are made salient to them, for instance by means of the messages sent as part of randomized controlled trials. Despite a series of significant and substantial results on tax collection, there is little evidence on the relative importance of these different mechanisms. Our contribution to the literature is to fill this knowledge gap about the mechanisms underlying taxpayers' reactions. The effects of the *audit-statistics* and *audit-endogeneity* messages are consistent with growing evidence that providing taxpayers with information about enforcement has positive effects on tax compliance in a variety of contexts: U.S. self-employment income (Slemrod et al. (2001)), wage income taxes in Denmark (Kleven et al. (2011)), individual public-TV fees in Austria (Fellner, Sausgruber, and Traxler (2013)), firms' VAT payments in Chile (Pomeranz (2015)), individual municipal taxes in Argentina (Castro and Scartascini (2015)), U.S. tax delinquents (Perez-Truglia and Troiano (2018)) and an individual church tax in Germany (Dwenger, Kleven, Rasul, and Rincke (2016)). The insignificant effects of our public-goods message is also consistent with evidence that moral suasion messages do not have a significant effect on tax compliance (Blumenthal, Christian, and Slemrod (2001); Fellner et al. (2013); Castro and Scartascini (2015); Dwenger et al. (2016); Perez-Truglia and Troiano (2018)), with the exception of Hallsworth, List, Metcalfe, and Vlaev (2017).

This paper is closely related to a group of studies testing the predictions of *A*-*S* in a laboratory setting. For example, Alm et al. (1992) conducted a laboratory experiment where undergraduate students play a tax evasion game. Subjects can hide income from the experimenter, but some subjects will be randomly selected to be audited and, if they are caught evading, they will have to pay a penalty. They show that tax compliance in the game increases significantly with audit and penalty rates, but these effects are economically small, and smaller than those predicted by optimizing behavior in the context of *A*-*S*. The laboratory experiment setting of Alm et al. (1992) and other similar studies that followed has a series of advantages, such as full control over the rules of the game and freedom in the selection of the model parameters. However, these results are also subject to some of the common criticisms of laboratory experiments, which are conducted in artificial contexts with low stakes and where subjects have no previous experience. We contribute to this literature by showing that the *A*-*S* model does not fare substantially better in a natural field experiment with high stakes and experienced subjects. To give an intuition of how these two contexts compare, the taxes paid in the laboratory experiments is typically less than USD10 per subject. In contrast, subjects in our field experiment paid an average of USD 11,800 in VAT and other taxes in the 12 months before our experiment alone.[5] Subjects in laboratory experiments are typically undergraduate students playing the tax game for the first time and with no prior experience paying taxes in the real world. In contrast, subjects in our field experiment are experienced firm owners who have been registered with the tax agency, and thus paying taxes, for an average of 15 years.

Our findings also contribute to the more general debate about the determinants of tax compliance. One of the main questions in this literature the relative contribution of tax authorities' enforcement efforts and taxpayers' tax morale – in other words, whether firms

[5]More specifically, an average of USD7,770 in VAT and USD407 in other taxes, as reported in Table B.1.

and individuals do not evade more taxes because they cannot do so, or because they do not want to (Luttmer and Singhal (2014)). There is strong evidence supporting the importance of the tax morale mechanism. In some contexts, for instance in the case of self-employed individuals, evasion is much lower than one would expect from a simple optimization exercise (e.g., Alm, McClelland, and Schulze (1992)). Our evidence indicates that this interpretation must be nuanced: taxpayers may fail to evade more not because they are unwilling to do so, but because they have not figured out what profit-maximizing behavior.

Finally, our paper is also part of small but growing literature on whether firms deviate from simple benchmarks of profit maximization (Romer (2006); DellaVigna and Gentzkow (2017)). We document two sources of profit-maximizing frictions. First, the fact that tax compliance is elastic with respect to the visibility of audit statistics but inelastic with respect to the audit probability and penalties rate suggests the presence of optimization frictions. Second, the fact that firms have biases in beliefs about audit probabilities suggests the presence of information frictions. More precisely, we show that, on average, firms perceive a probability of being audited that is almost three times as high as the actual one, but have unbiased beliefs about penalty rates. While this over-estimation of audit probabilities had been reported before,[6] this is, to the best of our knowledge, the first evidence of misperceptions in a high-stakes context.

The paper is organized as follows. Section 2 discusses the experimental design. Section 3 presents the econometric specifications and hypotheses. Section 4 presents the data sources and discusses the implementation of the field experiment. Section 5 presents the results on the average effect of the audit message, while Section 6 present the different tests of the underlying mechanisms. The final section concludes.

2 Experimental Design

2.1 *Baseline* Letter

Our experiment consisted of a mailing campaign from Uruguay's IRS, which included a number of treatment and subtreatment arms. Rather than comparing firms that received a letter to firms that did not, all of our analysis is based on comparisons between firms that received letters with subtle variations in their content. We can thus net out the potential

[6]This finding of overestimation of audit probabilities is consistent with prior survey evidence (Harris and Associates (1988); Erard and Feinstein (1994); Scholz and Pinney (1995)). However, this evidence is elicited from wage-earners, for whom the misperception of audit probabilities is mostly inconsequential due to widespread third-party reporting (Kleven et al. (2011)). In our context of small and medium enterprises, the financial stakes of audit probability misperceptions can be substantial.

effects of simply receiving a letter from the tax authority, which might induce – for instance – a reminder to pay due taxes.

These letters consisted of a single sheet of paper with the name of the recipient in the header, the official letterhead of the IRS, and the scanned signature of the IRS General Director. These letters were folded and placed in a sealed envelope with the official letterhead of the IRS in the outside. They were sent by certified mail, which guarantees that the letters were delivered directly to the recipient, who must sign upon receipt.

The first type of letter is the *baseline* letter, a sample of which is provided in Appendix A.1. The *baseline* letter contained some information about the goals and responsibilities of the tax authority, which the IRS routinely includes in its communications with firms. The text explained that the individual was randomly selected to receive this information, that the letter was for information only, and there was no need to reply or to provide any documentation to the IRS. The letters in the other treatment arms included the same text as in the *baseline* letter, but also included an additional paragraphs, printed in a larger type size and in boldface.

2.2 ***Audit-Statistics*** Letter

In the *audit-statistics* letter type, we provided firms with information about the audit and penalty rates. According to the Allingham and Sandmo (1972) model, we would expect risk-averse firms to be interested in this information because it would help them optimize their evasion decisions and potentially increase their bottom line.[7] Furthermore, this information would seem to be particularly valuable in the context of limited information about audits. For instance, it is easy to find information online about factors potentially relevant for firms' decision-making, such as the inflation rate or exchange rates. However, it is extremely difficult to find any information about audit probabilities and actual penalties paid by evading firms — tax authorities seem to prefer to conceal this information.

Appendix A.2 presents a sample of the *audit-statistics* letter type. Compared to the *baseline* letter, it contained the following additional paragraph with information about the audit probability (p) and penalty rates (θ) for a random sample of firms similar to that of the recipient:

> "On the basis of historical information on similar businesses, there is a probability of *[p%]*that the tax returns you filed for this year will be audited in at least one of the coming three years. If, pursuant to that auditing, it is determined that tax

[7]We assume that firms in our sample are risk averse, which is plausible since we deal mainly with small and medium firms. However, the A-S model has been generalized to settings with risk-neutral agents.

> evasion has occurred, you will be required to pay not only the amount previously unpaid, but also a fee of approximately *[θ%]* of that amount."

Note that we communicate the probability that firms will be audited in at least one of the three following years because IRS experts stated that this was the relevant probability for firms' decision-making: Uruguay's tax law indicates that tax audits should cover the previous three years of tax returns and, as a result, the probability that the current year's tax report will be audited is roughly equal to the probability that the firm gets audited at least once over the following three years.

In our sample, the average value of p is 11.7%, while the average value of θ is 30.6%. Tax agencies in most countries do not publish data on the values of p and θ, which makes it difficult to compare the Uruguayan case to other contexts. In the United States, for which some comparable data are available, these two parameters are on the same order of magnitude: self-employed individuals face a p of 11.42% and a base θ of 20%.[8]

The goal of this treatment arm is to generate exogenous variation in the firms' perceptions about audit probabilities and penalty rates. Because of legal and other constraints, we could not assign different firms to different sets of information about these factors. We induced non-deceptive exogenous variation in messages that may affect these perceptions by exploiting the sampling variation in statistics about audits and penalties.

More specifically, we divided the firms into five groups of "similar firms," corresponding to the five quintiles of total sales revenues. For each firm we then drew a random sample of 50 other firms from the same quintile (i.e., "similar firms"), from which we computed the averages of p and θ. This randomization strategy led us to 940 different combinations of p and θ. These estimates of p and θ were unbiased and consistent with the explanation given in the footnote — the information provided to the recipients was thus nondeceptive. We included a footnote in the letter to detail how we had estimated the values of p and θ:

> "Estimates are based on data from the 2011–2013 period for a group of firms with similar characteristics, for instance, in terms of total revenue. The probability of being audited was calculated as a percentage of audited firms in a random sub-sample of firms. The rate of the fee was estimated as an average of a random sub-sample of audits."

The values of p range from 2% to 25%, with an average of about 11.7%. The values of θ range

[8]First, there is an annual probability of being audited of 2.1%, according to the ratio of returns examined for businesses with no income tax credit and with a reported income between 25,000 and 200,000 USD (Table 9a of IRS, 2014). Each audit covers the previous 3 to 6 years, which implies that the the probability that the current year's tax filing will be eventually audited ranges from 5.88% to 11.42%. Second, IRS usually imposes a basic penalty of θ=20%, although the penalties can be higher in more severe cases.

from 15% to 68%, with an average of about 30.6%. Figure 1 presents the audit probability and penalty size distribution by quintile of firms sales, and the distribution of the generated parameters within group. Each row corresponds to a different quintiles of total sales – we denominate these groups 1 through 5. The vertical line denotes the average probability based on all members of the group. If we based our estimates of p and θ included in the letter on the population of firms, every member of the group would have received the same signal (the vertical line). Since we computed p and θ from samples of 50 firms, sampling variation implies that different members of each group received different signals. For example, Figure 1.a.1. shows that in group 1 the average p over all group members is 8.1%, while the histogram depict the different signals actually sent to firms within the group. These signals are centered around the average p, but they range from 2.5% to 20%. Note that there are some differences in the vertical lines across groups: for firms in the first quintile, the average randomized p is 8.1%, and this value increases monotonically up to 13.4% for those in the top quintile. This means that a small share of the variation across subjects in the values of p and θ included in the letter results from non-random variation across groups, while the within-group variation is fully due to random variation – this is an important factor for the econometric model presented below.[9]

2.3 *Audit-Threat* Letter

To complement the evidence from the *audit-statistics* sub-treatments, we implemented an alternative way of randomizing perceptions about audit probabilities. We devised a treatment arm called *audit-threat* letter that randomly assigned firms to groups with different probabilities of being audited, with a certain probability in the following year. A sample of the audit-threat letter is presented in A.3. The *audit-threat* letters were identical to the *baseline* letter, except for the following additional paragraph:

> "We would like to inform you that the business you represent is one of a group of firms pre-selected for auditing in 2016. A *[X%]* of the firms in that group will then be randomly selected for auditing."

We devised this treatment arm with the IRS audit department, which specified a group of high-risk firms. We randomly assigned them to two groups. The audit department committed to carry out audits on 25% of the firms in one of the groups and 50% of the firms in the other

[9] To measure the share of the variation that corresponds to the cluster size, we regress each parameter on the quintiles of sales revenues. Regressing p over VAT sales quintiles dummies results in $R^2 = 0.118$, while regressing θ over the same dummies results in $R^2 = 0.007$.

group in the following year. These were the two randomly assigned probabilities of audit (X=25% and X=50%) that we communicated in our letters in this treatment arm.

This *audit-threat* treatment arm was applied to a different sample of high-risk firms from that of the rest of the experiment. These different experimental samples imply that the effects of the *audit-threat* treatment arm cannot be directly compared to those of the *baseline* letter. We use this treatment arm as a separate auxiliary experiment.

2.4 *Audit-Endogeneity* Letter

The *audit-statistics* and *audit-threat* treatment arms conveyed quantitative information about audit probabilities and penalty rates. We also wanted to incorporate to our research design a message related to the qualitative nature of the tax administration's audit process.

Most tax agencies, including Uruguay's, take into account firm characteristics when deciding which ones to audit. They assign higher audit probabilities to firms that are considered to have a higher risk of evading. As a result, evading taxes typically increases the probability of being audited. In the canonical model of tax evasion, the audit probability is exogenous, but several authors, such as Allingham and Sandmo (1972); Yitzhaki (1987); Andreoni, Erard, and Feinstein (1998); Slemrod and Yitzhaki (2002), introduce variations in the model in which the audit probabilities are determined endogenously. In the context of these models, if unsuspecting firms learn about the endogenous nature of the audit probability, they should revise their tax evasion decisions and reduce the amount of tax evaded.[10]

We thus use this insight from economic theory to devise a message for firms about the nature of the audit process. We label this the *audit-endogeneity* message. We asked our counterparts at the IRS to use their evasion risk scoring to divide a small sample of firms into two groups - those suspected of evading taxes and those not suspected of evading at taxes. We then computed the difference in 2011–2013 audit rates between the two groups: the rates were approximately twice as high for the latter group. We used this information to create the message in the *audit-endogeneity* letter type, which adds to the *baseline* letter the following paragraph (see sample in Appendix A.4):

> "The IRS uses data on thousands of taxpayers to detect firms that may be evading taxes; most of its audits are aimed at those firms. Evading taxes, then, doubles your chances of being audited."

As in the *audit-statistics* treatment arm, two mechanisms could be at play with this *audit-*

[10]Konrad, Lohse, and Qari (2016) present suggestive evidence of this mechanism in the context of a laboratory experiment: taxpayers facing a situation where suspicious attitudes toward tax officers increase the probability of being audited increase their tax compliance by 80%.

endogeneity letter. On the one hand, tax compliance may increase when firms internalize the fact that audits are endogenous to their evasion behavior—the importance of their own actions becomes more palpable. On the other hand, firms may re-optimize their tax compliance decisions incorporating the information we provided about how audits change as a function of their behavior. If, on average, recipients revised their beliefs about the degree of endogeneity of audits upward (downward), the latter mechanism should increase (decrease) their tax compliance.

2.5 ***Public-Goods*** Letter

We also devised a treatment arm to provide a benchmark for the effect of messages intended to increase tax compliance, but that was not directly related to audits. In line with previous studies (see e.g., Blumenthal et al. (2001); Fellner et al. (2013); Pomeranz (2015); Dwenger et al. (2016)), we chose to include a message about public goods that could boost tax morale and thus increase compliance.

In coordination with IRS staff and authorities, we designed a non-pecuniary message to increase compliance that we expected could increase compliance. The message provided information about the cost of evasion in terms of the provision public goods, in the spirit of the model of Cowell and Gordon (1988).[11] The *public-goods* letter is identical to the *baseline* letter, with the exception of the addition of a specific paragraph. The paragraph lists a series of services that the government could provide if tax evaders reduced their evasion by 10% (see Appendix A.5 for a sample of the letter):

> "If those who currently evade their tax obligations were to evade 10% less, the additional revenue collected would enable all of the following: to supply 42,000 portable computers to school children; to build 4 high schools, 9 elementary schools, and 2 technical schools; to acquire 80 patrol cars and to hire 500 police officers; to add 87,000 hours of medical attention by doctors at public hospitals; to hire 660 teachers; to build 1,000 public housing units ($50m^2$ per unit). There would be resources left over to reduce the tax burden. The tax behavior of each of us has direct effects on the lives of us all."

We used estimates from different governmental agencies to design this message.[12]

[11] This message is also related to the laboratory experiment from Alm et al. (1992), which presents evidence that one of the reasons why people decide to pay taxes is their valuation of the public goods provided by means of the tax revenues.

[12] These agencies were: Administracion Nacional de Educacion Publica (ANEP), CEIBAL, Ministerio de Salud Publica (MSP), Ministerio del Interior (MI), Ministerio de Vivienda, Ordenamiento Territorial y Medio Ambiente (MVOTMA).

2.6 Survey Data

Finally, to disentangle the mechanisms behind the different messages, we designed a survey to be conducted with a sample of owners from our main subject pool. The IRS, with the support of the Inter-American Center of Tax Administrations and the United Nations, had previously administered a survey on the costs of tax compliance for small and medium-sized businesses. We collaborated with the tax authority in the design and implementation of this survey. The survey included a specific module to address the issues in our experimental design. More specifically, we included questions to elicit the degree to which our information provision experiment affected recipients' beliefs about audit probabilities and penalty rates. The survey also included seven additional modules designed by the IRS about the costs of tax compliance and other topics.

Appendix A.6 shows a sample of the email with the invitation to participate in the online survey. We partnered with local and international universities to increase respondent confidence and to highlight that the survey was part of a scientific study, and not an audit or compliance exercise by the IRS. To further ensure trustworthy responses, the IRS assured potential respondents that the survey responses would remain anonymous, and they could not be traced back to specific individuals or firms. To measure the effect of our experiment on the beliefs elicited by means of the survey, we embedded a code in the emailed survey link to identify which treatment arm of the experiment the recipient was assigned to (i.e., which of the four letter types and, within the *audit-statistics* treatment, which combination of p and θ). These codes did not uniquely identify any single firm, but they allowed us to link treatment arms and survey responses while still maintaining the anonymity of responses.

Appendix A.7 shows an excerpt from the survey instrument corresponding to our survey module. These questions were designed to assess whether the audit-statistics message shifted perceptions of the letter recipients, by means of the two following questions:

> Perceived Audit Probability: "In your opinion, what is the probability that the tax returns filed by a company like yours will be audited at least in one of the next three years (from 0% to 100%)?"
>
> Perceived Penalty Rate: "Let us imagine that a company like yours is audited and that tax evasion is detected. What, in your opinion, is the penalty (in %) as determined by law that the firm must pay in addition to the originally unpaid amount? For example, a fee of X% means that, for each $100 not paid, the firm would have to pay those original $100 plus $X in penalties."

After each of these questions, we elicit how certain the subject is about his or her response in a 1 to 4 scale from "not at all sure" (1) to "very sure" (4). For the sake of completeness, we also included a question in the survey intended to measure the subject's awareness about the endogeneity of audit probabilities—this question and its results are described in detail in Appendix B.6.

3 Econometric Specifications and Hypotheses

3.1 Average Effect of Audit Message

The experimental design relies on the comparison of outcomes between treatment arms, and in the case of the *audit-statistics* and *audit-threat* letters, comparisons across sub-treatment arms. For the comparison between treatment arms, consider the sample of firms assigned to either *baseline* letter or one of the other letter types, indexed by j*: audit-statistics*, *audit-endogeneity* or *public-goods*. The baseline specification is given by the following regression:

$$Y_i = \alpha + \beta \cdot D_i^j + X_i \delta + \epsilon_i \tag{1}$$

The outcome variable (Y_i) is the total outcome in the 12-month post-treatment period (i.e., after the delivery of the treatment letter). D_i^j is a dummy variable that takes the value 0 if i was assigned to the *baseline* letter, and the value 1 if i was assigned to letter type j. Finally, X_i is a vector of control variables. When outcomes are persistent over time, which applies to the case of VAT payments, the use of pre-treatment controls can help reduce the variance of the error term and thus results in gains in statistical power (McKenzie (2012)). Specifically, our baseline specification includes the outcomes during each of the previous 12 pre-treatment months as control variables (X_i).

We rely on a Poisson regression model to estimate our baseline specification for two reasons. First and foremost, the Poisson model allows effects to be proportional – indeed, the coefficients can be readily interpreted as semi-elasticities.[13] Second, the Poisson model accounts for the bunching at zero of the dependent variable. In any case, we present robustness checks with alternative regression models, including OLS and Tobit models. We also estimate separately the effects on the extensive margin. In the interest of transparency, for each coefficient related to post-treatment effects, we also present a falsification test based on pre-treatment "effects"; that is, we re-estimate the model, but with pre-treatment outcomes as the dependent variable. We should expect these pre-treatment effects to be close to zero and statistically insignificant.

[13]The Poisson model is based on the following specification: $\log(Y_X) = \alpha + \beta X + \varepsilon$. The effect of a unit change in X can be re-expressed in log-units of the dependent variable, $\beta = \log(Y_{X=x+1}) - \log(Y_{X=x})$, and thus can be approximated as a percent-change: $\beta = \log(Y_{X=x+1}) - \log(Y_{X=x}) \approx \dfrac{Y_{X=x+1} - Y_{X=x}}{Y_{X=x}}$.

This baseline specification allows us to test our first set of hypotheses: whether providing letters with information about enforcement increases tax compliance. The β coefficient for the *audit-statistics* model reflects the impact of providing information about audits. The findings in the previous literature indicate that this should have a positive effect on tax collection. The β coefficient for the *audit-endogeneity* regression, in turn, provides a benchmark for the *audit-statistics* results by capturing the effect of providing other type of information related to audits (i.e., not probabilities of audits nor penalty rates). Finally, the *public-goods* regression is a further benchmark, in this case providing the effect of conveying information not related to audits.

3.2 Heterogeneity by Prior Beliefs

The second econometric model is also based on a hypothesis derived from A-S: firms with different prior beliefs about the probability of being audited should react differently to information about this probability. We construct a proxy for prior beliefs for a particular firm based on two parameters: the number of years for which the firm has been registered with the tax agency, and the number of years that it has been audited during that period. The intuition behind this proxy is that there is little publicly available information about audit probabilities, so firms form their priors based on their own audit experience. For instance, when firms register with the tax authority, its prior may follow the beta distribution with parameters α_0, β_0. The firm has been registered for T_i years before our experiment. During that period, the firm has had $N_i \leq T_i$ yearly audits. The firm'"s posterior belief should thus follow a beta distribution with parameters $\alpha_1 = \alpha_0 + N_i,\ \beta_1 = \beta_0 + T_i - N_i$. The mean posterior belief for firm i will thus be $\hat{p}_i = \frac{\alpha_0 + N_i}{\alpha_0 + \beta_0 + T_i}$. To generate these prior proxies, we set $\alpha_0 = 0.3, \beta_0 = 1$ because these values result in an average posterior that matches the actual average audit probability in our subject pool.

The regression specification that captures these heterogeneous results is give by the following equation:

$$Y_i = \alpha + \rho \cdot I_{\{\hat{p}_i \leq 0.117\}} + \beta_1 \cdot D_i^j \cdot I_{\{\hat{p}_i \leq 0.117\}} + \beta_2 \cdot D_i^j \cdot I_{\{\hat{p}_i > 0.117\}} + X_i\delta + \epsilon_i \qquad (2)$$

where $I_{\{\hat{p}_i \leq 0.117\}}$ is a dummy variable equal to one if firm i's prior belief is smaller than 11.7%, the average of the audit probabilities conveyed in our *audit-statistics* treatment letters, and zero otherwise. The coefficient β_1 for the *audit-statistics* treatment thus corresponds to the effect of this treatment for subjects with prior beliefs about audit probabilities below the average message. Following A-S, we should expect these individuals to revise their beliefs upwards and thus increase their tax payments (i.e., $\beta_1 > 0$). On the contrary, β_2 captures the treatment effect for subjects with priors above the average message in the letters. This information should induce on average a downward revision in their perceived probabilities, which should lead to a reduction in their payments (i.e., $\beta_2 < 0$).

Finally, we report in Appendix B.3 the results under different calibrations of $\{\alpha_0, \beta_0\}$, which remain qualitatively and quantitatively similar to those obtained with $\{\alpha_0 = 0.3, \beta_0 = 1\}$.

3.3 Heterogeneity by Sub-Treatments

A third econometric model allows us to measure the effect of messages with different levels audit probabilities and penalty rates. Only the *audit-statistics* letters contained variation in these parameters. We estimate the following regression for this group:

$$Y_i = \alpha + \gamma_p \cdot p_i + \gamma_\theta \cdot \theta_i + \sum_{g=2}^{5} \pi_g \cdot I_{\{i \in g\}} + X_i \delta + \epsilon_i \tag{3}$$

where $p_i \in (0,1)$ is the signal about the audit probability included in the letter sent to subject i, and $\theta_i \in (0,1)$ is the penalty rate included in the letter sent to the same firm. The $I_{\{i \in g\}}$ variables are four dummies to control for the heterogeneity in the five groups of similar firms from which we drew the sample to calculate p_i and θ_i. Including these controls ensures that we only exploit the exogenous variation in p_i and θ_i induced by our experimental design – that is, the heterogeneity due to sampling variation. The scale of the right hand side variable and the use of a Poisson regression model imply that the resulting coefficients can be directly interpreted as elasticities. For instance, an estimate of $\gamma_p = 1$ would imply that a 1 percentage point increase in the audit probability conveyed in the letters increased VAT payments by 1%. From *A-S*, we can expect that $\gamma_p > 0$ and $\gamma_\theta > 0$ – i.e., firms tax payments are increasing in the perceived probability of audit and evasion penalty rates. Moreover, we can compare the values of these regression estimates to the predictions from the callibration of the *A-S* model.

A similar econometric model can be used for firms assigned to the *audit-threat* letter:

$$Y_i = \alpha + \gamma_p \cdot p_i + X_i \delta + \epsilon_i \tag{4}$$

where $p_i \in \{0.25, 0.50\}$ is the audit probability included in the *audit-threat* letter sent to firm i. As in the previous model, from *A-S* we expect a positive estimate of γ_p.

4 Data Sources and Implementation of the Field Experiment

4.1 Institutional Context

Uruguay is a South American country with an annual GDP per capita of about USD 15,000 in 2015. Total tax revenues (i.e., for all levels of government) were about 19% of GDP in 2015 and, as it is common in many countries, VAT represents the largest source of tax revenue, accounting for roughly 50% of the total tax revenues.[14] Firms are required to remit VAT payments all along the production and distribution chain.[15] The standard VAT rate is 22%, and a small number of specific products (a basket of basic foodstuff) either have a 10% rate or are exempt from the tax.

Although we study the impact of our experimental interventions on other taxes, our main focus is on the VAT, which represents the largest tax liability for firms in Uruguay. An important factor in our context is that VAT has limited third party reporting. The amounts reported for other sources of taxable income are subject to automatic third party reporting, and thus the taxpayers would be caught evading even without audits. For instance, wage income is hard to conceal in modern economies because employers are usually required to report their employees' earnings to the tax authority. As a result, tax evasion can be detected and deterred even without conducting an audit (Kleven et al. (2011)). Even if VAT requires a paper trail, this reporting has significant limitations in practice. For once, the paper trail breaks down when reaching the consumer. Moreover, firms can collude with each other, or with customers, to tamper with the paper trail, for instance by offering discounts for sales without receipts (Pomeranz (2015); Naritomi (2016)). At the time of our experiment, firms the IRS had not implemented standardized electronic receipts, which could facilitate and automatize the cross-check of the VAT trail to detect evasion. In this institutional setting and in the case of VAT liabilities for small and medium firms, audits represent the main tool for evasion deterrence – they are the primary mechanism through which tax administrations can detect VAT evasion Gomez-Sabaini and Jimenez (2012); Bergman and Nevarez (2006), and thus firms should pay significant attention to them.

[14]Own calculations based on data from the Central Bank of Uruguay and from the Internal Revenue Service. Other sources of tax revenues include the personal income tax, the corporate tax, and some specific taxes to consumption, businesses and wealth.

[15]Firms may credit VAT paid on input costs (i.e., imports and purchases from their suppliers) against the total sales of goods and services to their costumers (i.e., "tax debit"). They pay VAT to the IRS only on the excess of the total "tax debit" over the tax credit. If the tax credit exceeds the debit, the excess may be carried over for future tax years. While the VAT should in theory be similar in its effects to a retail sales tax, in practice the two types of taxes differ in some substantial aspects (Slemrod (2008)).

Tax morale in Uruguay — the intrinsic motivation to comply with taxation — is believed to be among the highest in Latin America, and possibly comparable to that in some developed nations. For instance, according to survey data from the 2010–2013 wave of the World Values Survey, 77.2% of respondents from Uruguay stated that evading taxes is "Never Justifiable," while this proportion is 68.2% on average (population-weighted) for all other Latin American countries, and 70.9% for the United States. Tax evasion is conversely relatively low in Uruguay. According to estimates from Gomez-Sabaini and Jimenez (2012), evasion of VAT in Uruguay was around 26% in 2008, the third lowest rate among the nine Latin American countries included in the study, and roughly comparable to the 22% evasion rate computed for Italy in 2006 (Gomez-Sabaini and Moran (2014)).[16]

4.2 Subject Pool and Randomization

Our experiment was conducted in collaboration with the IRS of Uruguay. As of May 2015, there were 120,142 firms registered in the agency's database. A subsample of 4,597 firms preselected by the IRS was put aside for the *audit-threat* sample, which we call the secondary experimental sample. Of the remaining firms, we followed a series of criteria to select our main experimental sample. We first excluded some firms by request of the IRS. For instance, we excluded firms subject to special regimes for VAT payments (very small or very large firms). We also only kept in the experimental sample firms that had made VAT payments in at least three different months in the previous 12-month period[17] and those with a total value added of at least USD 1,000. [18]

To maximize the impact of our information provision experiment, we did our best to ensure that the letters would be delivered to the firms' owners, or at least to the individuals making the day-to-day decisions. In some cases, owners provide the address of external accountants instead of their own or their firms'. We removed from the sample firms that were registered with an accountant's mailing address as their own (the IRS keeps records of addresses for all registered accountants). Moreover, in very large firms the effect of the information could be substantially diluted, since it would probably not reach the owner or the individuals making decisions about tax compliance. For that reason, we also excluded from our subject pool firms with a total value added above USD 100,000 during the previous

[16]Gomez-Sabaini and Jimenez (2012)computes those rates by applying the "indirect" method to estimate tax evasion. This method is based on the comparison of collected VAT to aggregate consumption data from the System of National Accounts (SNA).

[17]The sample selection was conducted in May 2015, so this 12-month period spans from April 2014 to March 2015.

[18]For the sake of simplicity, all amounts shown in this paper are adjusted by inflation and converted from Uruguayan pesos to U.S. dollars using the nominal exchange rate from August 2015

12 months.

This series of criteria left us with a subject pool of 20,471 firms for the main experimental sample. All these firms were randomly assigned to receive one the four letter types, with the following distribution: 62.5% to the main treatment arm, the *audit-statistics* letter, and 12.5% for each of the three remaining letter types (*baseline*, *audit-endogeneity*, and *public-goods*).[19] After removing the roughly 18.5% of the letters that were returned by the postal service, the final distribution of letter types was 10,272 to *audit-statistics*; 2,064 to *baseline*, 2,039 to *audit-endogeneity*, and 2,017 to *public-goods* (total N = 16,392).

The 4,597 firms in the secondary sample were assigned to receive the *audit-threat* letter. Half of them were randomly assigned to the message of a 25% audit probability, and the other half to the 50% audit probability. After excluding the 12% of letters returned by the postal service, we were left with 2,015 firms in the 25% probability group and 2,033 firms in the 50% probability group (total N = 4,048).

Columns (1) through (4) of Table 1 allows us to compare the balance of pre-treatment characteristics between firms assigned to the different letter types in the main experimental sample. The firm characteristics in the table include VAT payments before the experiment, the age of the firm and the number of employees, among others. Additionally, for each characteristic, column (5) presents the p-value of the test of the null hypothesis that the averages for these characteristics are the same across all four letter types. As expected, the differences across letter types are economically unimportant and not statistically significant. Columns (6) through (8) of Table 1 present a similar balance test, except for the secondary sample used for the *audit-threat* arm. Again, the characteristics are balanced across firms that received the 25% threat letter and firms that received the 50% threat letter.

Table 2 in turn provides some descriptive statistics for the firms in our subject pool. Column (2) corresponds to all firms in the main experimental sample. On average, these firms had paid USD 1,890 in VAT over the previous three months (implying a value added of about USD 8,600), they had been registered with the IRS for 15.3 years, they had an average of 4.8 employees, 14% of them had been audited at least once over the previous three years, and 22% belonged to the retail sector.

Column (1) of Table 2 corresponds to the universe of all registered firms. By design, firms in our experimental sample are smaller, both in terms of number of employees and in the level of VAT payments. Finally, column (3) of Table 2 provides statistics about the secondary experimental sample (i.e., for the *audit-threat* treatment arm). While some statistically significant differences exist between firms in the two groups, they are still broadly

[19]The randomization to letter types was stratified by the quintiles of the distribution of value added over the 12 months previous to the randomization.

comparable in size. The main difference between firms in the two experimental samples is that the audit rates were 9 percentage points higher in the *audit-threat* sample. This difference is by design, because the IRS selected firms classified as high-risk for this treatment arm, and these had a higher propensity to be targeted for audits in the past.

4.3 Outcomes of Interest

The letters were provided to Uruguay's postal service on August 21, 2015. The vast majority of the letters were delivered during September, and therefore we set August as the last month of the pre-treatment period, and October as the first month of the post-treatment period. The main outcome of interest in our study is the total amount of VAT liabilities remitted by taxpayers in the 12 months after receiving the letter.[20] To test for the persistent of our treatment effects, we define a second period of observation between October 2016 and September 2017, i.e., up to two years after the intervention.

On average, the total amount of VAT paid by firms that received the *baseline* letter in the 12-month pre-treatment period is about USD 7,700, while the amount for the corresponding post-treatment period is approximately USD 6,500. This negative trend in VAT payments can be explained by the fact that this sample contains a high share small firms, which have a high turnover rate. The size of post-treatment VAT payments varied substantially, ranging from a 10th percentile of USD 400 to a 90th percentile of USD 16,550. [21]

Furthermore, we can break down firms' VAT payments according to their timing. We can observe the date of the transfer to the IRS as well as the month for which the payment was imputed. Firms can back-date payments, and transfer funds to cover liabilities from previous periods. Since firms typically make VAT payments on a monthly basis, they normally cover the current and the previous month, which we call concurrent payments (73.5% of the firms in July 2015). We classify payments for two or more months in the past as retroactive payments (3.9% of the firms in July 2015).

Finally, although we focus on VAT payments in our analysis, we obtained data from the IRS on the other main taxes paid by the firms, which include corporate income taxes and net worth taxes. These two taxes and the VAT jointly represent more than 96% of the total tax burden of firms. We use payments of these taxes as additional outcomes of interest, especially to study whether firms effectively change their overall compliance, or if they substitute evasion of VAT for that of other taxes.

[20] This variable includes VAT payments and also VAT withholding made by third party agents. For simplicity, we refer to this total as "VAT payments".

[21] Table B.1 in Appendix B.1 presents detailed descriptive statistics about the distribution of pre- and post- treatment payments for firms that received the *baseline* letter type.

4.4 Survey Implementation

The IRS communicates mainly by postal mail, and it thus has mailing addresses for all registered firms, but it only keeps records of email addresses for a subset of firms that used their online services in the past. We sent invitations to all the firms in the main experimental sample with a valid email address. The email invitations to the online survey were sent by email on May 2016, about nine months after our mailing experiment. The last column of Table 2 presents the average characteristics among the 3,845 firms that were invited to the survey. The firms invited to the survey were very similar in characteristics to those in the main experimental sample (column 2).

Our purpose was to elicit the beliefs of firm owners. We did not include email addresses that in the full sample were repeated more than three times, which most likely corresponded to accountancy firms representing multiple small and medium enterprises. Even after applying this criteria, the IRS records could not ensure that the registered email address corresponded to the firms' owners. We thus asked the survey respondent to self-identify as one of the following five types: owner, internal accountant, external accountant, manager, or other employee. From the 3,845 recipients that we invited to participate in the survey, we received 2,331 responses (response rate of 60.6%). Of these 2,331, 45% self-identified as owner, 4.4% as internal accountant, 5.4% as external accountant, 1.9% as manager, 4.5% as other employee, and the remaining 38.8% did not provide a response to this question. For the purpose of our estimates we use only respondents who self-identified as owners.

Finally, the IRS requested that none of the questions were mandatory. Among respondents self-identified as owners, the missing data rate for our three key questions is between 19.5% and 23.3%, which is comparable to the average non-response rate for all questions in the survey (17.8%).

5 Results: Average Effect of Audit Messages

5.1 Main results

Our baseline results capture the impact of our mailing campaign on our outcomes of interest (subsequent tax payments). These results are obtained by comparing the post-treatment tax payments of firms assigned to the *audit-statistics*, *audit-endogeneity*, and *public-goods* letters with the payments from recipients of the *baseline* letter. We interpret these as the effects of the corresponding messages (*audit-statistics*, *audit-endogeneity*, and *public-goods*). We start by describing the effects of our main experiment, the *audit-statistics* message, and discuss the benchmark results from the other two sub-treatments in the following subsection.

Figure 2 summarizes these baseline results. In each of the three panels, we plot the difference between the VAT payments of the firms assigned to the *baseline* letter to the payments of firms assigned to each of the other three treatment arms for the two quarters preceding our mailing campaign and for the eight subsequent quarters.[22] The effects of each of the treatment arms are computed by means of Poisson regressions, so that the coefficients can be directly interpreted as semi-elasticities. These are simple regressions, with no additional control variables.

Figure 2.a shows that the *audit-statistics* message had statistically and economically significant effects on VAT payments. For instance, the point coefficient corresponding to the third post-treatment quarter implies that the *audit-statistics* message increased the VAT payments by 8.7% ($p = 0.012$) [23]. The effects are similar in magnitude for all the post-treatment quarters during the first year (e.g, 7.8%, 5.8%, 8,7% and 5.7% in the first through fourth quarter). During the second year, the effects start vanishing over time. The coefficients on the pre-treatment VAT payments correspond to a falsification test for our experiment. As expected by the random assignment of firms to different treatment arms, the pre-treatment differences in VAT payments are economically small and not statistically significant (effects of less than 1% and about -1.7%, statistically insignificant in both cases).

Table 3 presents the baseline regression results. These estimates are obtained by means of Poisson regressions with pre-treatment controls, following the econometric specification from Section 3.1. The first column presents our baseline results: the effects of each of the three treatments (*audit-statistics* in Panel A, *audit-endogeneity* in Panel B, and *public-goods* in Panel C) compared to the outcomes for firms that received the *baseline* letter. The post-treatment coefficients correspond to regressions with total VAT paid in the 12 months after the delivery of the letter as the dependent variable (October 2015 - September 2016). Additionally, as placebo tests, the pre-treatment coefficients correspond to regressions with total VAT paid in the 12 months prior the delivery of the letter as the dependent variable (September 2014 - August 2015).

The first column from Table 3 corresponds to average effects of each treatment arm for the entire sample. The post-treatment coefficient of *audit-statistics* (panel a.) indicates that firms receiving this message paid 6.3% more VAT in the 12 months after the intervention on average. This effect is not only highly statistically significant ($p = 0.013$), it is also economically substantial. Using the estimated average evasion rate of 26% from Gomez-

[22]We top-coded all outcomes of interest at 99.99% to avoid the contamination of the results by typos, measurement error and outliers.

[23]Because the order of magnitude of the estimated coefficients is small enough, we can accurately approximate the percentage change in the dependent variable simply using the log-difference that results from the Poisson regression. We discuss the estimated coefficient in these terms.

Sabaini and Jimenez (2012), the effect amounts to a reduction in the evasion rate of 24% $(= \frac{6.3\%}{26\%})$.

The (placebo) effect on pre-treatment outcomes is close to zero (-0.8%), not statistically significant at standard levels, and even more precisely estimated than the corresponding post-treatment effect (the standard error on the pre-treatment coefficient is 0.021, 16% smaller than the corresponding standard error of 0.025 for the post-treatment coefficient).

The effects of our *audit-statistics* treatment are not directly comparable to those of the audit message from Pomeranz (2015) because the messages differed in content and because the two studies cover firms from different countries and with different characteristics. Nevertheless, Table 4 from Pomeranz (2015) indicates that the deterrence letter in that study led to an increase in VAT payments of 7.6%, which is similar in magnitude and statistically indistinguishable from the 6.3% effect of our *audit-statistics* message. Moreover, our results are consistent with a broader literature that finds effects of messages about enforcement on tax compliance in a variety of contexts: self-employed income in the United States (Slemrod et al. (2001)), wage income taxes in Denmark (Kleven et al. (2011)), individual public-TV fees in Austria (Fellner et al. (2013)), individual municipal taxes in Argentina (Castro and Scartascini (2015)), U.S. tax delinquencies in the United States (Perez-Truglia and Troiano (2018)) and an individual church tax in Germany (Dwenger et al. (2016)).

The second column from Table 3 replicates the analysis for the second year after the treatment, i.e. October 2016 - September 2017. For this period, the effect of the treatment is less than half than that of the first year, and it is not statistically significant, even though the precision of the estimate is similar in the two time periods. This is consistent with the pattern of effects by quarter depicted in Figure 2.a, which show that the effects in the second year fall substantially when compared to those of the first year. These results are not unexpected. On the one hand, over time, individuals may forget the information conveyed by the letter, or it may become less salient. Individuals may also update their beliefs and perceptions, for instance because of new events such as audits and information campaigns. On the other hand, this is consistent with previous evidence on the effects of tax enforcement messages. Figure 2 in Pomeranz (2015) shows that the effects of the main information campaign in that study are also substantially higher in the first 12 months, and fall substantially and become virtually zero in the 18th month (the last plotted result in the Figure).

Table 3 also presents results complementary outcomes. As discussed in the previous section, firms in Uruguay make payments for their current liabilities, but also for taxes that correspond to previous periods—either because they owe past taxes, or because they are revise their accounts and correct past mistakes, or because they impute invoices that were not available at the time of the original payment. When firms that engage in tax evasion face an increased threat of being audited, we can expect them to increase their tax payments (reduce their evasion) in the future, but we can also expect them to retroactively revise their payments for previous time periods to reduce or eliminate their past evasion. We explore this possibility with the results presented in columns (3) and (4) of Table 3, which split the effects of our treatment arms on retroactive and concurrent payments. These results correspond to the first year after the treatment. The messages about audits had an economically and statistically significant effect on retroactive payments. Indeed, because of the much lower baseline rate, the effects on retroactive payments are larger in magnitude than the effects on the concurrent payments. For instance, the effect for *audit-statistics* message is 38.1% (p = 0.004) for the retroactive payments and 4.4% (p = 0.087) for the concurrent payments.

We have so far established that firms in the *audits-statistics* treatment arms increased their VAT payments compared to recipients of the *baseline* letter. Our analysis focuses on VAT liabilities, which represent the largest fraction of tax payments by firms in our sample and because VAT is the tax that leaves more room for exploiting loopholes and because of the lack of (or at least limited) third-party reporting and ensuing cross-checks. However, our mailing referred to taxes in general and did not mention VAT nor any other specific tax. In fact, the effects on VAT may not represent a net increase in tax payments; firms may increase their evasion (i.e., reduce their payments) of other taxes they are liable for.

The results in columns (5) and (6) of Table 3 shed light on these issues. The columns present the effects of our treatment arms for different taxes: (5) other taxes (mostly the corporate income tax), and (6) total (i.e., VAT + other). The evidence suggests that, far from crowding out other tax payments, the *audit-statistics* had positive effect on non-VAT payments. Indeed, the effects on payments of other taxes are as economically and statistically as significant as those on VAT payments. The *audit-statistics* had an effect of 6.3% on VAT payments, while the effect on other tax payments was 7.7%. (p = 0.038). This difference implies that the effect of the *audit-statistics* message on other tax payments was 22.2% larger compared with VAT payments, and that our treatment increased overall tax payments, with no payment or evasion shifting.

5.2 Benchmarks to the ***audit-statistics*** message

The previous subsection addressed the question of whether the post-treatment tax payments of firms assigned to the *audit-statistics* letters were higher than those of recipients of *baseline* letters. Additionally, our research design allows us to compare the effects of the *audit-statistics* message to the effects of the other messages, some related to audits (*audit-endogeneity*) and some unrelated (*public-goods*).

Figure 2.b shows that the *audit-endogeneity* treatment arm also induced a significant change in VAT payments compared to the *baseline* letter. The quarterly effect of adding the *audit-endogeneity* message is positive and statistically significant in the first post-treatment year, and ranges between 10% and 12.4%. These effects are slightly higher than those of the *audit-statistics* treatment arm, but these differences are not statistically significant at standard levels. As in the case of the *audit-statistics* message, the effects of the *audit-endogeneity* are persistent over the first year, but start to vanish during the second post-treatment year. The results of the falsification test are also similar to the ones observed in the *audit-statistics* treatment: the differences for the two pre-treatment quarters are economically small and not statistically significant for the *audit-endogeneity* treatment arm. This overall pattern o results is confirmed by the regression results presented in Panel b. from Table 3. The coefficient in the first column indicates that the *audit-endogeneity* message increased subsequent VAT payments by 7.4% (p = 0.021), with the placebo experiment confirming the pre-treatment balance in the outcome of interest. This effect is similar in magnitude to the 6.3% effect of our *audit-statistics* message (difference not statistically significant). The results observed for the second year effect are also consistent with the ones of the *audit-statistics* message: second year effects are smaller in magnitude than first year effects and statistically insignificant. The *audit-endogeneity* message also had an economically and statistically significant effect on both retroactive (31.4%) and concurrent payments (6.1%), as in the case of the *audits-statistics* message. Finally, the results in columns (5) and (6) of Table 3 indicate that, also as in the case of the *audit-statistics* treatment, far from crowding out other tax payments, the *audit-endogeneity* messages had a statistically significant 8.4% effect on non-VAT tax payments.

Figure 2.c, in turn, depicts the effects of the *public-goods* treatment arm. Adding this message to the *baseline* letter did not seem to have an effect as high and stable as that of the other two treatment arms neither on the first nor the second year. However, while there are large differences in post-treatment VAT payments, inspection of the event-study reveals substantial differences in the pre-treatment period between firms assigned to the *public-goods* letters and those assigned to the *baseline* letter. The results on balance between groups indicate that this pre-treatment difference for one of the sub-treatments is most likely due to chance. A proper assessment of the effect of the *public-goods* message requires us to control for these differences in pre-treatment outcomes, and this is presented in Panel c. from Table 3. The placebo regression that pools the payments from the 12 months immediately preceding the treatment indicates that this outcome was balanced, and that the results presented in the Figure for the two pre-treatment quarters are thus due to chance. However, the effect for the first year of the post-treatment period for the *public-goods* message is positive at 4.3%, and

falls to 0.6% in the second year. Neither of them are statistically significant at standard levels (p = 0.147 and p=0.879 respectively). Compared to the other two treatment arms, these effects are smaller, but due to the lack of precision they are statistically indistinguishable. The result for the first year, in particular, is not a precisely estimated zero. Regarding the other variations of the outcome, the *public-goods* message did not have a statistically significant effect either on retroactive payments or on current payments (p = 0.146 and p = 0.304, respectively). Finally, the effect of the *public-goods* message on other tax payments (column 5 and 6 in Table 3) is close to zero (0.1%) and not statistically significant at standard levels. The effect of the *public-goods* message on total tax payments was also small (1.8%) and statistically insignificant. Contrary to the effects of the two messages about audits, the effects of the *public-goods* message is not robust across specification, and it is close to zero for some of the variations of the outcome that we consider (those for the second year, those for other taxes, etc.). This finding of insignificant effects of a public goods-type of message is consistent with the robust finding that moral suasion messages do not have a significant effect on tax compliance (Blumenthal et al. (2001); Fellner et al. (2013); Castro and Scartascini (2015); Perez-Truglia and Troiano (2018); Dwenger et al. (2016), with the exception of Hallsworth et al. (2017).

Appendix B.2 presents some additional robustness checks, including alternative specifications based on OLS, Tobit and Probit models, as well as focusing on the extensive margin and using an independent source of administrative data. We show that the main findings are qualitatively and quantitatively similar, and thus robust to these alternatives.

6 Results: Mechanisms

To sum up, the results presented so far are broadly consistent with the evidence in the previous literature: the information about audits (especially the *audit-statistics* message) significantly increased tax compliance. This section addresses the question of whether these effects are consistent with the *A-S* model: each subsection presents a discussion and a test of the potential underlying mechanisms.

6.1 First Test: Survey Data

According to the Allingham and Sandmo (1972) model, the *audit-statistics* message could have a positive effect on tax compliance if it increased the expected value of the evasion fines – i.e., by increasing either the perceived probability of being audited or the perceived fine (or both). We explore this hypothesis based on data from our post-treatment survey.

Our survey data consists of 365 firms in the *audit-statistics* group and 137 in the pooled control group.[24]

Figure 3.a and 3.b depict the distributions of perceptions about audit probabilities and penalty rates, respectively, as elicited from the survey. The shaded gray bars show the distribution of perceptions for individuals from firms that received the *baseline* letter or the *public-goods* letter (that is, those in the pooled control group). The red dashed curve corresponds to the distribution of signals sent to the firms in the *audit-statistics* letters. The comparison between the shaded bars and the red curve from Figure 3.a suggests that, on average, respondents in the control group substantially overestimated the probability of being audited. While our statistics indicate a probability of about 11.7%, the mean perception for the control group is 40.7% ($p < 0.01$ for the difference). Conversely, the comparison between the shaded bars and the red curve from Figure 3.b suggests that respondents had unbiased beliefs regarding the penalty rates: the average penalty is 30.7%, while the mean of the perceived penalty is 30.5% in the control group. Moreover, respondents were mostly confident on their responses: only 16.2% of those in the control group reported being "Not sure at all" about their perceived probability of audit, and only 18.1% stated the same for their perception of the penalty rate. Our respondents thus reported being confident even though their estimates were substantially different from the actual values of the parameters – in fact, the average perceived probability for those in the control group reporting to be "Very confident" in their response was 42.1%, substantially higher than the actual 11.7%.

This positive bias in the perceived audit probability can be explained by the availability heuristic bias (Kahneman and Tversky (1974)). According to this model, individuals judge the probability of an event by how easily they recall instances of it. Even though audits are rare, the fact that they may be visible for colleagues and even sometimes salient in the media may induce firms to perceive them to be more frequent than they actually are. Indeed, there is evidence that individuals overestimate the probabilities of a wide range of rare events of a similar nature (Lichtenstein, Slovic, Fischho, Layman, and Combs (1978); Kahneman, Slovic, and Tversky (1982)).

Regarding the distribution of beliefs, the systematic bias in the perceptions of audit probabilities and penalty rates in the control group is not consistent with the predictions of the *A-S* model. The *audit-statistics* letters, on average, provided information about audit probabilities that were substantially higher than those perceived by the experimental subjects.

[24]To increase the statistical power, we defined a control sub-sample group by pooling subjects from the *baseline* and the *public-goods* groups since both received messages with no specific information about audit probabilities or fines. The resulting pooled control group has 137 observations – Appendix B.5 shows that the results are similar, but less precisely estimated, when we only use recipients of the *baseline* letter for the control group.

If this information reduced this systematic bias, it should have lowered the perceived audit probability, which in the A-S framework should result in reduced tax compliance. The results from the previous section, however, should substantial positive effects on tax compliance from the *audit-statistics* letters.

Our post-treatment survey allows us to test this hypothesis more directly. From the assignment of firms to the different treatment groups, we can measure the effect of the *audit-statistics* letter on the perceived p and θ. The shallow bars in Figure 3 depict the distribution of perceptions for respondents from firms in the *audit-statistics* treatment arm compared to respondents in the treatment group. Inspection of Figure 3.a indicates that the *audit-statistics* message did indeed reduce the perceived probability of being audited, from an average of 40.7% in control group to an average of 35.2% in the *audit-statistics* group (p = 0.03). Meanwhile, Figure 3.b shows that the *audit-statistics* message had a small effect on the perceived penalty rate, decreasing it from an average of 30.5% for the control group to an average of 29.9% for the *audit-statistics* group, although this difference is not statistically significant (p-value = 0.85). This analysis, while compelling, only provides a lower bound for the effect of our letters on perceptions. While we are confident that the our certified letters reached firms' owners, or at least someone in charge of running the business, we cannot be as confident about whether the same person received the email invitation to complete the survey. Moreover, the survey invitation was circulated 9 months after our mailing, and it is feasible that recipients had started to forget the messages we included at that point.

To sum up, the evidence suggests that the *audit-statistics* treatment arm may have reduced the average perceived audit probability. This evidence is inconsistent with the A-S model, according to which the *audit-statistics* message, by reducinng perceived audit probabilities, should have reduced VAT payments instead of increasing them. Consistent with this evidence, Appendix B.6 suggests that the effect of the *audit-endogeneity* message is not due to an update of recipients' beliefs about the degree of endogeneity of audits, because recipients were already well aware of the endogeneity of audit probabilities.

While survey data has some known limitations, these do not seem to be driving our results. One of the most relevant challenges is that in some contexts where individuals have high uncertainty, they report probabilities of 50%, which might be a substitute for expressing their uncertainty (Bruin, Fischbeck, Stiber, and Fischhoff (2002); Bruine de Bruin and Carman (2012)). In our survey, 33.3% of responses about the perceived audit probability and 16.4% of responses about the penalty rate are exactly 50% for those in the control group. However, there are at least two reasons why this may not be a problem in this context. First, we can use the data on self-reported certainty about the response, in a 1-4 scale. Again in the control group, the average certainty for perceived audit probability is 2.18 for individuals who responded 50% and 2.56 for individuals who responded a different value. For the responses about the perceived penalty rate, the average certainty is 2.18 for individuals who responded 50% and 2.48 for those who responded another value. In other words, the 50% responses do not seem to be used to express uncertainty. Second, we can also conduct our analysis ignoring 50% responses and treating them as missing data. Even with this extreme criteria, individuals in the control group still substantially over-estimate the probability of being audited: these respondents report an average perception of 31.5%, compared to the actual probability of 11,7%.

Another potential concern with our survey results is that our subjects may be confused about the correct meaning of an "audit" in this context. We can test this directly with survey data. About 8.9% of survey respondents report to have been audited in the previous three years, which is very close to the average audit probability of 11.7% in our administrative data.

Finally, our survey results may be more trustworthy than those of internet surveys based on random mailings because we specifically targeted business owners, who should be familiar with fractions and probabilities. At the very least, they need some rudimentary arithmetics and understanding of percentages to compute the VAT and other tax liabilities.

Given these limitations, our survey results might only be suggestive. In what follows, we provide two alternative tests of the same hypothesis that do not rely on survey data.

6.2 Second Test: Heterogeneity with Respect to Prior Beliefs

According to the Allingham and Sandmo (1972) model, another mechanism at play in our context is that the *audit-statistics* message should have a larger positive effect on tax compliance for individuals with a lower prior belief about audit probabilities.

We discussed our setup for estimating these heterogeneous effects in Section 3.2 above. The results of these heterogeneity tests are presented in Table 4. Column (1) presents the results for the entire sample, which are identical to the main results in column (1) of Table 3. In turn, columns (2) through (5) present the effects for different subgroups of the population.

The results in columns (2) and (3) break down the effects of the messages by whether the recipient's prior belief ($\hat{p}$) is below or above the average audit probability included in the audit-statistics message (p=11.7%). The results indicate that the effects of the *audit-statistics* letter are somewhat larger for firms with relatively low $\hat{p}$ (6.9%) compared to those with relatively high $\hat{p}$ (4.9%) – the difference between the two coefficients, however, is not statistically significant. This evidence is not consistent with the prediction of the *A-S* model, according to which the effect of the *audit-statistics* message on tax compliance, which on average conveyed a signal of p=11.7%, should have been negative for the group with $p \leqslant \hat{11}.7\%$ and positive for the group with high $\hat{p}$ (>11.7%), since the signal would increase the perceived probability for the former and decrease it for the latter. The point estimates suggest that, if anything, the heterogeneous effects follow the opposite pattern, with higher positive effects for the group with low $\hat{p}$.

Appendix B.3 shows that these results are robust to a number of checks. For instance, the results are similar if we calibrate $\hat{p}$ to be centered at the average perceived audit probability instead of the actual audit probability; and the results are also similar if we compare recipients whose prior belief is above or below the specific audit probability signal included in the respective letter.

To illustrate that our results are not sensitive to the calibration of the Bayesian model, columns (4) and (5) of Table 4 provide a less parametric version of the test: we compute the heterogeneous treatment effects by considering separately firms that were audited in the previous 15 years (23,8% of the sample) and those that were not audited during that period (the remaining 76,2% of the sample). We use the previous 15 years because that is how far back the available IRS administrative records reach. The intuition is that firms that have been exposed to audits in the past should have higher prior belief about the audit probability than firms that have never been exposed to audits, and thus should react less (possibly with a negative sign) to the audit message. We find a pattern of heterogeneous in the "right" (i.e., the A-S) direction, but the difference in treatment effects for the two groups is small and not statistically significant: the effect is 4.7% for group the group of firms previously audited, and 7.1% for those with no prior audit experience (p-value of difference = 0.252).

As complementary evidence, we can use the survey data to conduct a validation exercise for the comparison between firms that were audited in the past and firms that were not. Among the 145 responses from the *baseline* and *public-good*s groups (our pooled control group for the survey), 10.3% of firms reported that they were audited in the past three years and the remaining 89.7% reported that they were not. Consistent with our assumption about prior beliefs, we find that respondents from firms that were audited in the recent past reported a higher average perceived probability of being audited (63.9%) than individuals from firms that were not recently audited (38.1%), a highly statistically significant difference (p-value<0.001). Moreover, this gap is economically significant: as reported in a companion paper (Bergolo, Ceni, Cruces, Giaccobasso, and Perez-Truglia (2018)), this indicator for having been recently audited is the single most important predictor of perceived audit probabilities among a host of different factors. The recent audit experience also has a positive effect on the perceived penalty rate, but this effect is economically and statistically less significant.[25]

[25]Respondents from firms that were audited recently report an average perceived penalty rate of 40.0%,

Finally, for completeness, the second and third panels of Table 4 present the results of heterogeneity for the remaining two treatment arms. For the *audit-endogeneity* message, if recipients changed their perceptions by increasing their perceived probability of audit, then the hypotheses about heterogeneous effects should be the same as for the *audit-statistics* treatment. Again, we do not find evidence of significant differences between those with higher and lower priors – if anything, the effect on tax compliance seems to be higher for those with a prior above the average audit probability. Regarding the *public-goods* group, we should not expect any specific pattern in the heterogeneity of results, and there does not seem to be a pattern at all: the four estimated coefficients are not statistically significant.

6.3 Third Test: Heterogeneity by Sub-Treatments

According to the A-S model, tax compliance should be increasing in both the audit probabilities p and the penalty rates θ. As a result, individuals who, by chance, receive a letter with a higher signal about any of these two parameters should have a stronger reaction – i.e., a higher treatment effect.

Figure 4.a provides an event-study analysis of the effect of the *audit-statistics* message (relative to the *baseline* letter) for two groups: firms that received letters with low signals of p (p<=12%) and those who received high signals of p (p>12%). The estimates in these figures are obtained from regressions that control for the group dummies, so they are identified purely by the random sampling variation in p. Note that the effects are extremely similar for the two groups. If anything, the treatment effects seem larger for the group with the low probability message. Figure 4.b presents the equivalent analysis for firms that received above- and below-median values of θ. Again, the effects are very similar for the two groups.

The variation we introduced in the p and θ conveyed in our *audits-statistics* letters allow us to conduct a more detailed analysis. Figure 5.a presents the effect of the audit-statistics message on VAT payments for each decile of the p included in the letter. Again, there is no clear pattern nor relationship between the effect of the *audit-statistics* message and the value of p included in the letter. Figure 5.b in turn shows the equivalent figure for the deciles of the conveyed θ instead. Again, there is no relationship between the effect of the *audit-statistics* message and the value of θ that was included in the letter.

Panel (a) in Table 5 presents the results from a more parametric analysis of these results. Column (1) of Table 5 presents estimates of the elasticities of VAT payments with respect to the conveyed p and θ in the *audit-statistics* subtreatments, as described in the econometric model of section 3.3. The estimates are close to zero, statistically insignificant and precisely

compared to 29.4% for resopndents from firms that were not audited recently. However, this differences is not statistically significant since the p-value of the difference test is 0.201.

estimated. The elasticity with respect to the audit probability in the first year after the treatment is 0.030 (SE of 0.236, p-value = 0.897). This means that increasing p by 1 percentage point would increase VAT payments by a mere 0.03%. The elasticity with respect to the penalty rate is −0.118 (SE 0.115, p = 0.304). These coefficients are precisely estimated, which means that we can rule out even moderate elasticities: the 90% confidence interval for the audit probability excludes elasticities above 0.418, and the 90% confidence interval for the penalty rate excludes elasticities above 0.071. It should be noted that the pre-treatment falsification test do not yield any statistically significant effect, and that the results are similar (no statistically significant elasticities) for the other specifications: for the second year, by timing of the payment (columns (3) and (4)), and by type of tax (columns (5) and (6)). In most cases, the estimates are not statistically significant. There are some exceptions like the coefficient associated with the audit probability for retroactive payments, but in this case the differences were statistically significant before the treatment and therefore this result is probably due to chance.

All together, this evidence suggests that firms did not react to values of the parameters contained in the message. Since we devoted a large fraction of our subject pool to this treatment arm, these elasticities are quite precisely estimated.

Our research design provides some complementary evidence. Qe can use the results from the *audit-threat* secondary sample, which generated alternative variation in the audit probabilities by randomly assigning firms to messages of audit probabilities of 25% and 50%. The pattern of results by quarter in Figure 4.c. and the regression results in panel (b) of Table 5 for this treatment arm broadly confirm the results for the *audit-statistics* treatment. There are no statistically significant differences in VAT payments for the two groups. While the *audit-threat* messages implies an elasticity of 0.376, which is borderline significant at the 10% level (p = 0.073), the pre-treatment (falsification) coefficient (-0.342) is also borderline statistically significant at the 10% level (p-value=0.055), indicating that the post-treatment effect implied by the coefficient may be spurious. In any case, this elasticity of 0.376 is still economically small. Appendix B.4 present a series of robustness checks (alternative specifications based on OLS, Tobit and Probit models, and using an alternative data source for the dependent variable), and the results are similar. An additional robustness check, presented in Table B.5 in Appendix B.4, shows that results are robust if, instead of estimating the elasticities with respect to p and θ separately, we estimate the elasticity to the expected fine $p * \theta$.

Finally, we can provide a more quantitative test of the null of the *A-S* model. While *A-S* predicts positive elasticities with respect to p and θ, the exact magnitudes of the predicted elasticities depend on the specific model at hand and on the values of the underlying parameters. We provide calibrations under a number of settings that have been considered in the literature, allowing for social preferences (Dhami and al Nowaihi (2007)), endogenous audit probabilities (Allingham and Sandmo (1972); Yitzhaki (1987); Andreoni et al. (1998); Slemrod and Yitzhaki (2002)), misperceptions about audit parameters (Alm et al. (1992)) and non-audit detection. We calibrate different combinations of these models to match the average VAT evasion rate in Uruguay (26%, as estimated by Gomez-Sabaini and Jimenez (2012)), and compute the elasticity of tax payments with respect to p and θ under each of the calibrations.

The details and results from these calibrations are presented in Appendix C. In our preferred calibration, we find an elasticity of tax payments with respect to the audit probability of 4.55 and an elasticity of tax payments with respect to the penalty rate of 3.48. Taking the main specification in column (1) of 5, we can test the null that the elasticity with respect to p and θ are equal to those in the preferred A-S calibration. We can reject the null that the elasticity is 4.55 for the audit probability, and that it is 3.48 for the penalty rate (both tests with p-values$<$0.001). These calibrated elasticities are also well beyond the confidence intervals for the coefficients estimated for the *audit-threat* model. And even though different calibrations can lead to somewhat different elasticities, the results would be similar under the alternative calibrations.

It should be noted that our discussion is based on the implicit assumption that a letter with a 1 percentage point higher signal will increase the perception of the parameter by the recipient by 1 percentage point. This is probably a strong assumption: some individuals may not have read the letter in its entirety, or they may have not entirely believed in the content of our message. If we instead assume that for every percentage point difference in the letter individuals only adjusted their beliefs by half a percentage point, then we should double the elasticities estimated in our regressions before comparing them to the calibrations of the A-S model. Even under this conservative assumption, we can still reject the null that the estimated and the calibrated elasticities of tax compliance with respect to the audit probability are equal ($p<0.001$). Indeed, even if assumed that for each percentage point difference in the letter individuals only adjusted their beliefs by one tenth of a percentage point, we would still fail to reject the null hypothesis that the estimated elasticities are equal to the A-S calibration.

These results on the elasticities for the *audit-statistics* and the *audit-threat* messages are related to recent papers such as Kleven et al. (2011) and Dwenger et al. (2016). In one treatment arm in their experiment, Kleven et al. (2011) show that randomizing employed individuals to a higher audit probability (100% instead of 50%) increases their tax compliance by an amount that is statistically significant but economically negligible. However, their findings do not constitute evidence against *A-S* because they conduct these experiments with wage earners, for whom evasion is almost always detected through automatic third-party reporting and without the need of audits. As a result, *A-S* would predict that this sample of taxpayers should not care about the probability of audits. In another experiment, Dwenger et al. (2016) show that announcing different probabilities of audits does not have a statistically significant effect on compliance with a small local church tax in Germany. However, because of statistical power, this study cannot rule out the presence of economically significant differences.

7 Conclusions

The canonical model of Allingham and Sandmo (1972) predicts that firms evade taxes by optimally trading off between the costs and the benefits of evasion. However, there is still no consensus about whether real-world firms react to audits in this way. We conducted a large scale field experiment in collaboration with Uruguay's tax authority to shed light on the factors behind firms' evasion behavior and their reactions to audits. We sent letters to 20,440 firms that collectively pay over 200 million dollars in taxes per year. We measure the effect of these signals on their subsequent perceptions about the auditing process, based on survey data, as well as on the actual taxes paid, according to administrative data.

The first part of our experimental design confirmed a now standard finding in the literature: providing letters with information about enforcement increases tax compliance. Do firms react optimally to this information? Our experimental design also tested the underlying mechanisms. First, we show that the information about audits in our letters actually decreased firms' perception of audit probabilities – in the A-S model, this should have decreased rather than increased compliance. Moreover, firms audited in the past report higher perceived probabilities of audits, but the effect of the letters on tax compliance is not higher for this group. Finally, our letters also provided exogenous yet non-deceptive signals about audit probabilities and penalty rates. The change in tax compliance was not a function of the provided information, even though the information significantly affected firms' perceptions.

Results from the three tests suggest that the average small and medium firms do not react to the audit messages in the calculating manner predicted by Allingham and Sandmo (1972). There are some caveats with this interpretation, of course. For example, the fact that the average firm does not behave as *A-S* predicts does not imply that none of the firms are behaving in this way. It is possible that some of the firms behave as *A-S* predicts, but our evidence suggest that those may not be the majority.

All in all, our evidence is not consistent with the predictions of Allingham and Sandmo

(1972). Audits may deter taxpayers from evading taxes in the same way that scarecrows deter birds from eating crops, which is more consistent with models of risk-as-feelings compared to alternative explanations such as salience and prospect theory.

Even though our experiment was designed to test the null hypotheses of Allingham and Sandmo (1972), we can still check whether our findings are consistent with some of these alternative models. We start with the salience model (Chetty et al. (2009)), according to which firms behave as if the the probability of detection and the penalty rate are zero unless these parameters are salient to them (for instance, by means of a letter like ours). A priori, salience seems to be a plausible explanation for why our audit messages increase tax compliance: when reminded about the values of the audit parameters, individuals stop behaving as if these were zero and start taking their real values into account for their optimizing decisions. However, the salience model fails to fit this and other findings. First and most important, by definition, salience models predict short-lived effects – for instance, if a consumer is reminded about a non-salient tax, that should affect the behavior of the consumer that day, but should not affect his consumption on, say, a month later. This prediction is then in sharp contrast with our evidence on the persistent effects of our audit messages, which seem to be present a year after the messages were transmitted. Second, the salience model is inconsistent with the evidence that individuals have the same reaction to the audit messages regardless of their prior beliefs, and regardless of the exact values of the signals included in the messages.

Another family of models that has been proposed to understand tax compliance are those of prospect theory Kahneman and Tversky (1979)) and loss aversion (Tversky and Kahneman (1992)). While these models may predict quantitatively different values, they share the same qualitative predictions of the A-S model on how audit messages should influence behavior as a function of prior beliefs and of the exact values included in the messages. Indeed, these models tend to predict that an intcrease in the tax rate should increase evasion (Dhami and al Nowaihi (2007)) .

Another plausible model to explain our findings is that of rational inattention (Reis (2006)), according to which firms may choose not to pay the attention to the audit parameters because it is not worth their time to do so. This explanation is inconsistent with the fact that this is a high-stakes decision: figuring out the optimal evasion rate could result in thousands of dollars of additional income for the owner per year, thereby requiring implausible large attention costs to rationalize this behavior.

It may also be argued that our subjects may have interpreted the reception of a letter such as ours as a signal that their firms were under the IRS radar. We were careful to mitigate this concern in our mailing. For instance, we highlighted the fact that the letter recipients were randomly selected, but individuals reading the messages may have ignored or overlooked this.

We also were very careful in selecting a non-threatening wording for both the *baseline* and the *audit-statistics* messages. But recipients of the *audit-statistics* letters may still believe that “firms like” theirs are audited with a given probability, but that their own might be audited with a higher probability because they received this letter. However,this interpretation is not consistent with two of the tests we implement in our analysis: even if simply receiving the letter has an effect on the expected probability of audit beyond the value of this parameter conveyed in the letter, we should still expect that this effect is a function of prior beliefs and of the content of the letter, which we do not find.

Finally, our paper is also part of small but growing literature on whether firms deviate from simple benchmarks of profit maximization (Romer (2006); DellaVigna and Gentzkow (2017)). Firm owners may lack the capacity to figure out optimal decisions.

This explanation, however, is at odds with the characteristics of our subject pool. As it has been documented in hundreds of studies, even individuals with no knowledge of finance seem to be able to make the type of simple trade-offs involving risk like the ones involved in the evasion “bet”. This should be even more true in our subject pool, individuals who are more educated than the average, and who have years of expertise in managing their business. Moreover, our subjects may have access to employees or accountants that can help them with their decision-making. However, we still document two sources of profit-maximizing frictions. First, the fact that tax compliance is elastic with respect to the visibility of audit statistics but inelastic with respect to the audit probability and penalties rate suggests the presence of optimization frictions. Second, the fact that firms have biases in beliefs about audit probabilities suggests the presence of information frictions. More precisely, we show that, on average, firms perceive a probability of being audited that is almost three times as high as the actual one, but have unbiased beliefs about penalty rates. While this over-estimation of audit probabilities had been reported before,[26] this is, to the best of our knowledge, the first evidence of misperceptions in a high-stakes context.

A final model that we consider is that of risk-as-feelings (Slovic, Finucane, Peters, and MacGregor (2004); Loewenstein et al. (2001)). A large literature in psychology provides evidence that, when individuals have to choose from options involving risk, even though they make mistakes, they make calculated choices with their prefronrtal cortex. However, when the risk choices have a negative emotional connotation (i.e., fear), the decisions are made with more primitive parts of the brain which are not involved in making calculations. This

[26] This finding of overestimation of audit probabilities is consistent with prior survey evidence (Harris and Associates (1988); Erard and Feinstein (1994); Scholz and Pinney (1995)). However, this evidence is elicited from wage-earners, for whom the misperception of audit probabilities is mostly inconsequential due to widespread third-party reporting (Kleven et al. (2011)). In our context of small and medium enterprises, the financial stakes of audit probability misperceptions can be substantial.

may result in choices and behavior that deviate from optimal decisions. This interpretation can fit our main findings. It can explain why the audit message affected tax compliance, yet nonetheless that effect was orthogonal to the specific information contained in the letter and orthogonal to the prior beliefs. Indeed, the fact that individual react to risky information neglecting the specific probabilities is widely documented in this literature, and known as probability neglect (Sunstein (2002)).

In sum, our preferred interpretation suggests that tax audits may deter taxpayers from evading taxes in the same way that scarecrows deter birds from eating the crops. However, we must note this is speculative, and further research is needed to test our hypothesis further.

Finally, we conclude with some policy implications for our findings. First, our survey data indicates that increasing transparency about audit probabilities could reduce the average perceived probability of being audited, with possibly negative effects on tax compliance. Indeed, the fact that information about auditing processes such as audit probabilities is not easily accessible online suggests that tax administrations may already be aware of this possibility, and should perhaps continue to limit the circulation of this information. Second, holding the actual detection rate fixed, tax agencies may benefit from making audits and other detection mechanisms more visible to taxpayers.[27] There is evidence that some tax agencies are already taking advantage of this approach. For instance, Blank and Levin (2010) finds a disproportionately large number of tax enforcement press releases during the weeks immediately preceding to Tax Day, presumably in order to scare taxpayers while they are preparing to file their annual tax returns. Third, our findings suggest that the tax compliance of small and medium firms are relatively large due to the presence of information and optimization frictions. If this is the case, tax agencies may leverage these frictions to increase compliance at a low cost.

[27]For a practical discussion on how to implement this type of policy, see for example Morse (2009). Furthermore, visibility can be used to improve compliance with other laws: for instance, Dur and Vollaard (2016) show experimental evidence that the use of a salience intervention can significantly reduce illegal garbage disposal.

References

Allingham, M. G. and A. Sandmo (1972). Income Tax Evasion: A Theoretical Analysis. *Journal of Public Economics 1*, 323–338.

Alm, J., B. Jackson, and M. McKee (1992). Estimating the Determinants of Taxpayer Compliance with Experimental Data. *National Tax Journal 45*(1), 107–114.

Alm, J., G. H. McClelland, and W. Schulze (1992). Why do people pay taxes? *Journal of Public Economics 48*(1), 21–38.

Andreoni, J., B. Erard, and J. Feinstein (1998). Tax Compliance. *Journal of Economic Literature 32*(2), 818–860.

Becker, G. S. (1968). Crime and Punishment: An Economic Approach. *Journal of Political Economy 76*(2), 169–217.

Bergman, M. and A. Nevarez (2006). Do Audits Enhance Compliance? An Empirical Assessment of VAT Enforcement. *National Tax Journal 59*(4), 817–832.

Bergolo, M., R. Ceni, G. Cruces, M. Giaccobasso, and R. Perez-Truglia (2018). Misperceptions about Tax Audits. *AEA Papers and Proceedings 108*, 83–87.

Blank, J. D. and D. Z. Levin (2010). When Is Tax Enforcement Publicized? Research Paper, Virginia Tax Review, 30; NYU Law and Economics.

Blumenthal, M., C. Christian, and J. Slemrod (2001). Do Normative Appeals Affect Tax Compliance? Evidence From a Controlled Experiment in Minnesota. *National Tax Journal 54*(1), 125–138.

Bruin, W. J. A. B. d., P. S. Fischbeck, N. A. Stiber, and B. Fischhoff (2002). What number is "fifty-fifty"? Redistributing excess 50% responses in risk perception studies. *Risk Analysis 22*(4), 725–735.

Bruine de Bruin, W. and K. G. Carman (2012). Measuring Risk Perceptions: What Does the Excessive Use of 50% Mean? *Medical Decision Making 32*(2), 232–236.

Castro, L. and C. Scartascini (2015). Tax Compliance and Enforcement in the Pampas Evidence From a Field Experiment. *Journal of Economic Behavior & Organization 116*, 65–82.

Chetty, R., A. Looney, and K. Kroft (2009). Salience and Taxation: Theory and Evidence. *The American Economic Review 99*(4), 1145–77.

Cowell, F. A. and J. P. F. Gordon (1988). Unwillingness to pay: Tax evasion and public good provision. *Journal of Public Economics 36*(3), 305–321.

DellaVigna, S. and M. Gentzkow (2017). Uniform Pricing in US Retail Chains. Working Paper 23996, National Bureau of Economic Research.

Dhami, S. and A. al Nowaihi (2007). Prospect theory versus expected utility theory: Why Do People Pay Taxes? *Journal of Economic Behavior and Organization 64*(1), 171–192.

Dwenger, N., H. Kleven, I. Rasul, and J. Rincke (2016). Extrinsic and Intrinsic Motivations for Tax Compliance: Evidence from a Field Experiment in Germany. *American Economic Journal: Economic Policy 8*(3), 203–232.

Erard, B. and J. S. Feinstein (1994). The Role of Moral Sentiment and Audit Perceptions in Tax Compliance. *Public Finance 49*, 70–89.

Fellner, G., R. Sausgruber, and C. Traxler (2013). Testing Enforcement Strategies in the Field: Threat, Moral Appeal and Social Information. *Journal of the European Economic Association 11*(3), 634–660.

Gomez-Sabaini, J. C. and J. P. Jimenez (2012). Tax structure and tax evasion in Latin America. *Macroeconomics of Development Series 118*.

Gomez-Sabaini, J. C. and D. Moran (2014). Tax policy in Latin America Assessment and guidelines for a second generation of reforms. *Macroeconomics of Development Series 133*.

Hallsworth, M., J. A. List, R. D. Metcalfe, and I. Vlaev (2017). The behavioralist as tax collector: Using natural field experiments to enhance tax compliance. *Journal of Public Economics 148*, 14–31.

Harris, L. and I. Associates (1988). 1987 taxpayer opinion survey. *Washington, DC: Internal Revenue Service Document*.

Kahneman, D., P. Slovic, and A. Tversky (Eds.) (1982). *Judgment under uncertainty: heuristics and biases*. Cambridge ; New York: Cambridge University Press.

Kahneman, D. and A. Tversky (1974). Judgment under Uncertainty: Heuristics and Biases. *Science 185*(4157), 1124–1131.

Kahneman, D. and A. Tversky (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica 47*(2), 263–291.

Kleven, H. J., M. B. Knudsen, T. Kreiner, S. Pedersen, and E. Saez (2011). Unwilling or Unable to Cheat? Evidence from a Randomized Tax Audit Experiment in Denmark. *Econometrica 79*(3), 651–692.

Konrad, K. A., T. Lohse, and S. Qari (2016). Compliance With Endogenous Audit Probabilities. *Scandinavian Journal of Economics*.

Lichtenstein, S., P. Slovic, B. Fischho, M. Layman, and B. Combs (1978). Judged frequency of lethal events. *Journal of experimental psychology: Human learning and memory 4*(6), 551.

Loewenstein, G. F., E. U. Weber, C. K. Hsee, and N. Welch (2001). Risk as feelings. *Psychological Bulletin 127*(2), 267–286.

Luttmer, E. F. P. and M. Singhal (2014). Tax Morale. *Journal of Economic Perspectives 28*(4), 149–168.

McKenzie, D. (2012). Beyond baseline and follow-up: The case for more T in experiments. *Journal of Development Economics 99*(2), 210–221.

Morse, S. C. (2009). Using Salience and Influence to Narrow the Tax Gap. *Loyola University Chicago Law Journal 40*, 483.

Naritomi, J. (2016). Consumers as Tax Auditors.

Perez-Truglia, R. and U. Troiano (2018). Shaming tax delinquents. *Journal of Public Economics 167*, 120–137.

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

Pomeranz, D. and J. Vila-Belda (2018). Taking State-Capacity Research to the Field: Insights from Collaborations with Tax Authorities.

Reis, R. (2006). Inattentive consumers. *Journal of monetary Economics 53*(8), 1761–1800.

Romer, D. (2006). Do Firms Maximize? Evidence from Professional Football. *Journal of Political Economy 114*(2), 340–365.

Scholz, J. T. and N. Pinney (1995). Duty, Fear, and Tax Compliance: The heuristic basis of citizenship behavior. *American Journal of Political Science 39*, 2.

Slemrod, J. (2008). Does It Matter Who Writes the Check to the Government? The Economics of Tax Remittance. *National Tax Journal 61*.

Slemrod, J. (2018). Tax Compliance and Enforcement. *Journal of Eonomic Literature Forthcoming*.

Slemrod, J., M. Blumenthal, and C. Christian (2001). Taxpayer Response to an Increased Probability of Audit: Evidence from a Controlled Experiment in Minnesota. *Journal of Public Economics 79*(3), 455–483.

Slemrod, J. and S. Yitzhaki (2002). *Chapter 22 - Tax Avoidance, Evasion, and Administration," in "Handbook of Public Economics" edited by: Auerbach, A. J. and Feldstein, M.* Elsevier.

Slovic, P., M. L. Finucane, E. Peters, and D. G. MacGregor (2004). Risk as analysis and risk as feelings: some thoughts about affect, reason, risk, and rationality. *Risk Analysis: An Official Publication of the Society for Risk Analysis 24*(2), 311–322.

Sunstein, C. (2002). Probability Neglect: Emotions, Worst Cases, and Law. *Yale Law Journal 112*(1), 61–107.

Tversky, A. and D. Kahneman (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty 5*(4), 297–323.

Yitzhaki, S. (1987). On the Excess Burden of Tax Evasion. *Public Finance Review 15*(2), 123–137.

Table 1: Balance of Observable Firm Characteristics across Treatment Groups

	Main Sample					Secondary Sample		
	Audit Statistics (1)	Public Goods (2)	Audit Endogeneity (3)	Baseline (4)	p-value test (5)	Audit Threat (25%) (6)	Audit Threat (50%) (7)	p-value test (8)
Share paid VAT taxes (3 months pre-mailing)	0.92	0.94	0.93	0.93	0.18	0.90	0.89	0.54
	(0.00)	(0.01)	(0.01)	(0.01)		(0.01)	(0.01)	
Amount of VAT paid (3 months pre-mailing)	1.87	1.96	1.93	1.91	0.56	1.74	1.75	0.95
	(0.03)	(0.07)	(0.07)	(0.06)		(0.10)	(0.09)	
Years registered in tax agency	15.34	14.75	15.70	15.01	0.27	19.45	19.42	0.94
	(0.17)	(0.22)	(0.54)	(0.22)		(0.28)	(0.29)	
Share audited between 2013-2015	0.11	0.10	0.09	0.10	0.30	0.13	0.15	0.38
	(0.00)	(0.01)	(0.01)	(0.01)		(0.01)	(0.01)	
Number of employees	4.81	4.66	4.88	5.09	0.96	4.83	4.88	0.80
	(0.26)	(0.54)	(0.57)	(0.64)		(0.13)	(0.12)	
Share retail trade sector	0.22	0.22	0.21	0.23	0.78	0.33	0.32	0.40
	(0.00)	(0.01)	(0.01)	(0.01)		(0.01)	(0.01)	
Share Agricultural, forest and others	0.03	0.03	0.03	0.03	0.77	0.03	0.04	0.12
	(0.00)	(0.00)	(0.00)	(0.00)		(0.00)	(0.00)	
Share construction sector	0.03	0.03	0.03	0.03	0.85	0.03	0.03	0.33
	(0.00)	(0.00)	(0.00)	(0.00)		(0.00)	(0.00)	
Share other sector	0.73	0.73	0.73	0.72	0.95	0.61	0.62	0.54
	(0.00)	(0.01)	(0.01)	(0.01)		(0.01)	(0.01)	
Observations	10,272	2,017	2,039	2,064		2,015	2,033	

Notes: Averages for different pre-treatment firm-level characteristics, by treatment group and type of sample. The main sample includes all firms selected as described in section 4.2. The secondary sample includes high risk firms selected by the IRS. Standard errors in parentheses. The last column of each sample reports the p-value of a test in which the null hypothesis is that the mean is equal for all the treatment groups. Data on VAT amount and firm characteristics comes from administrative tax records (including monthly payments, annual tax returns and auditing registers).

Table 2: Comparison of Firm Characteristics for All Firms, for Firms in Main Sample, for Firms in the Secondary Sample and for Firms Invited to the Survey

		Experimental Sample		
	All firms (1)	Main (2)	Secondary (3)	Invited to the survey (4)
Share paid VAT taxes (3 months pre-mailing)	0.78	0.93	0.89	0.93
	(0.42)	(0.26)	(0.31)	(0.26)
Amount of VAT paid (3 months pre-mailing)	3.72	1.89	1.74	1.89
	(11.55)	(2.83)	(4.25)	(2.98)
Years registered in tax agency	14.21	15.26	19.44	14.46
	(14.85)	(17.16)	(12.84)	(10.08)
Share audited between 2013-2015	0.06	0.10	0.14	0.08
	(0.32)	(0.40)	(0.46)	(0.35)
Number of employees	12.65	4.84	4.89	6.43
	(302.97)	(26.60)	(5.76)	(53.77)
Share retail trade sector	0.13	0.22	0.33	0.15
	(0.34)	(0.41)	(0.47)	(0.36)
Share Agricultural, forest and others	0.03	0.03	0.03	0.02
	(0.17)	(0.17)	(0.18)	(0.15)
Share construction sector	0.03	0.03	0.03	0.03
	(0.17)	(0.17)	(0.17)	(0.18)
Share other sector	0.84	0.73	0.61	0.80
	(0.37)	(0.45)	(0.49)	(0.40)
N	120,142	16,392	4,048	3,845

Notes: Column (1) includes all firms that remitted at least one payment in 2014 or 2015. Column (2) includes the subset of firms selected for the experimental sample according to the criteria described in section 4.2. Column (3) represents a group of high risk firms that were selected from a special sample defined by the IRS and received the *audit-threat* letter. Column (4) corresponds to firms with valid e-mail addresses on file with the IRS, and therefore selected to participate in the on-line survey conducted after the experiment. All data is based on administrative tax records (monthly payments, annual tax returns and auditing registers).

Table 3: Average Effects of *Audit-Statistics*, *Audit-Endogeneity* and *Public-Goods* Messages: Time Horizon, VAT Payment Timing and Other Taxes

	By Time Horizon		By Payment Timing		By Tax Type	
	First Year (1)	Second Year (2)	Retroactive (3)	Concurrent (4)	Non - VAT (5)	VAT + Non-VAT (6)
a. Audit - Statitstics (N= 10,272) vs Baseline (N= 2,064)						
Post-Treatment	0.063**	0.032	0.381***	0.044*	0.077**	0.052**
	(0.025)	(0.031)	(0.131)	(0.026)	(0.037)	(0.026)
Pre-Treatment	-0.008	-0.003	-0.132	-0.005	0.016	0.037
	(0.021)	(0.025)	(0.105)	(0.022)	(0.045)	(0.026)
b. Audit - Endogeneity (N= 2,039) vs Baseline (N= 2,064)						
Post-Treatment	0.074**	0.038	0.314**	0.061*	0.084**	0.079***
	(0.032)	(0.039)	(0.139)	(0.033)	(0.041)	(0.031)
Pre-Treatment	-0.006	0.072**	0.008	-0.003	0.006	0.027
	(0.027)	(0.030)	(0.125)	(0.027)	(0.056)	(0.031)
c. Public - Goods (N= 2,017) vs Baseline (N= 2,064)						
Post-Treatment	0.043	0.006	0.195	0.032	0.001	0.018
	(0.030)	(0.036)	(0.134)	(0.031)	(0.048)	(0.029)
Pre-Treatment	-0.004	0.031	-0.192	0.002	-0.047	-0.004
	(0.026)	(0.028)	(0.124)	(0.026)	(0.046)	(0.027)

Notes: * significant at the 10% level, ** at the 5% level, *** at the 1% level. Robust standard errors. The results are based on a Poisson regression, so coefficients can be interpreted directly as semi-elasticities. Panel a. compares the *audit-statistics* message with the *baseline* letter while panels b. and c. replicate the analysis for *audit-endogeneity* and *public-goods* messages. In the first row of each panel, the dependent variable is the amount of VAT payments (in dollars) after receiving the letter. The second row presents a falsification test in which we estimate the same regression but using the amount contributed before receiving the mailing (pre-treatment) as the dependent variable. All regressions are estimated with a set of monthly controls that correspond to the year before the outcome, i.e. in the post treatment outcome we include monthly VAT payment controls from September, 2014 to August, 2015 and in pre-treatment outcome we include the same variables for the September, 2013 - August 2014 period. We also restrict the analysis to those firms that effectively received the letter. Columns (1) and (2) report the effect of treatment by time horizon. Column (1) shows the first year effect estimates (October 2015 - September 2016) while Column (2) reports the effect of the letter in the second year after the treatment (October 2016 - September 2017). Columns (3) and (4) show the first-year effect of treatment on retroactive (3) and concurrent (4) and VAT payments. Columns (5) and (6) report the first-year results by type of tax. Column (5) shows the effect of the treatment in other non-VAT tax payments while column (6) reports the effect on the total amount of taxes paid by the firms in the same period.

Table 4: Average Effects of *Audit-Statistics*, *Audit-Endogeneity* and *Public-Goods* Messages

		By $\hat{p}$		Audited in 2001-2015	
	All	$\leq 11.7\%$	$> 11.7\%$	Yes	No
	(1)	(2)	(3)	(4)	(5)
a. Audit - Statistics vs Baseline					
Post-Treatment	0.063**	0.069**	0.049	0.047	0.071**
	(0.025)	(0.035)	(0.035)	(0.042)	(0.031)
Pre-Treatment	-0.008	-0.010	-0.005	-0.062*	0.016
	(0.021)	(0.028)	(0.029)	(0.037)	(0.025)
Observations	12,336	6,523	5,813	2,933	9,403
b. Audit - Endogeneity vs Baseline					
Post-Treatment	0.074**	0.063	0.097**	0.081	0.076**
	(0.032)	(0.045)	(0.043)	(0.055)	(0.039)
Pre-Treatment	-0.006	-0.010	-0.026	-0.054	0.008
	(0.027)	(0.036)	(0.036)	(0.047)	(0.031)
Observations	4,103	2,175	1,928	969	3,134
c. Public - Goods vs Baseline					
Post-Treatment	0.043	-0.015	0.066	0.070	0.032
	(0.030)	(0.039)	(0.041)	(0.053)	(0.036)
Pre-Treatment	-0.004	0.025	-0.018	-0.033	0.005
	(0.026)	(0.035)	(0.034)	(0.046)	(0.031)
Observations	4,081	2,091	1,990	983	3,098

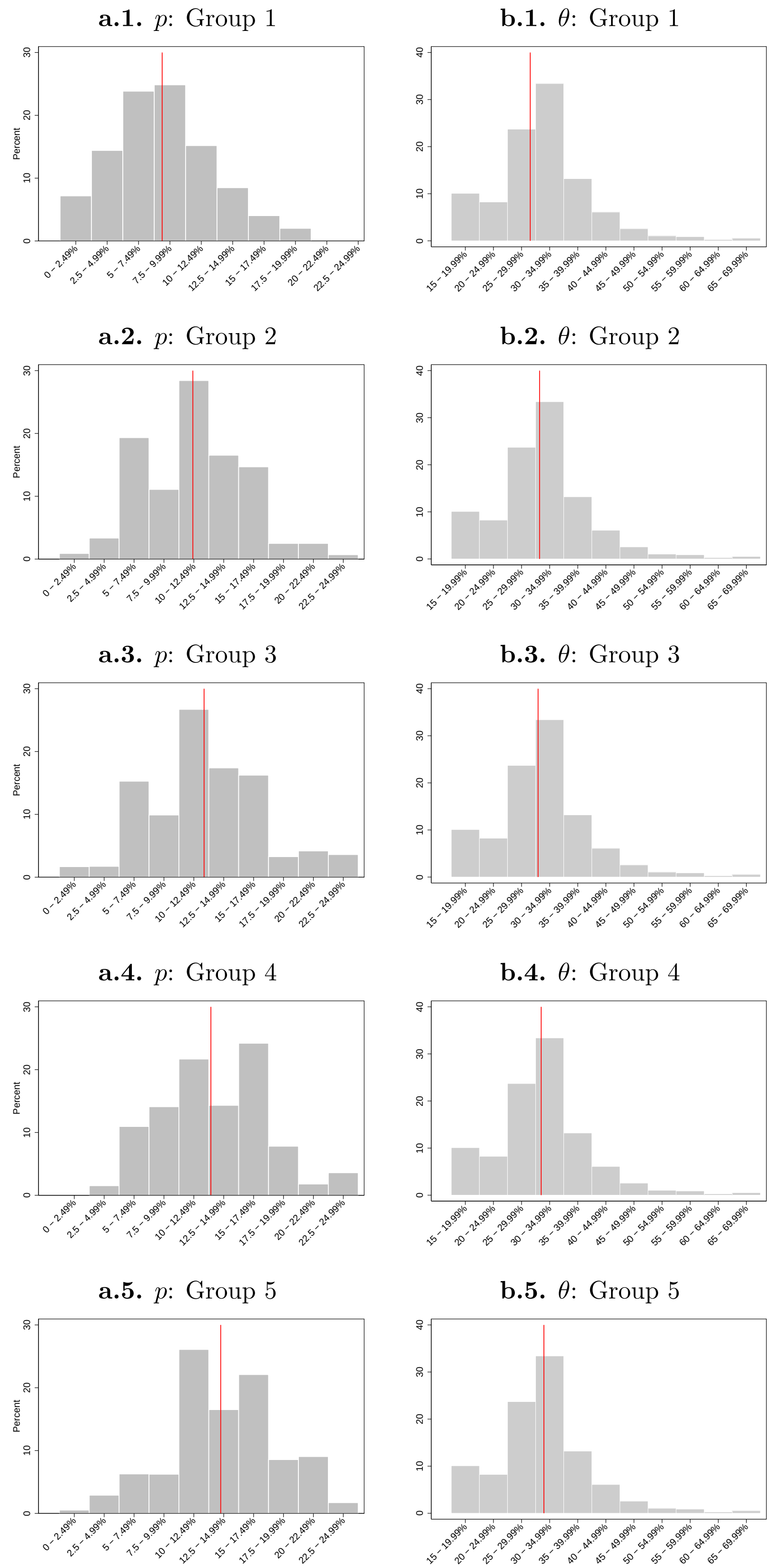
Notes: * significant at the 10% level, ** at the 5% level, *** at the 1% level. Robust standard errors. The results are based on a The results are based on a Poisson regression, so coefficients can be interpreted directly as semi-elasticities. Panel a. compares the *audit-statistics* message with the baseline letter, while panels b. and c. replicate the analysis for *audit-endogeneity* and *public-goods* messages. In the first row of each panel the dependent variable is the amount of VAT payments (in dollars) in the first year after receiving the letter (October 2015 - September 2016). The second row presents a falsification test in which we estimate the same regression, but using the amount contributed the year before receiving the mailing as the dependent variable (September 2014 - August 2015). All regressions are estimated with a set of monthly controls that correspond to the year before the outcome, i.e. in the post treatment outcome we include monthly VAT payment controls from September, 2014 to August, 2015 and in pre-treatment outcome we include the same variables for the September, 2013 - August 2014 period. We also restrict the analysis to firms that effectively received the letter as reported by the postal service. Column (1) reports the average effect of the treatment in the first year after receiving the letter for all firms treated. Column (2) presents the first-year effect of each letter on firms whose prior about the probability of being audited ($\hat{p}$) is greater than the mean value of the randomized p (11.7%). Column (3) does the same for firms with a $\hat{p}$ below 11.7%. Column (4) reports estimates for firms that were audited at least once between 2001 and 2015, and Column (5) for the group of firms that were not. .

Table 5: Elasticities of Evasion with Respect to Audit Probability and Penalty Rate, *Audit-Statistics* and *Audit-Threat* Sub-Treatments

	By Time Horizon		By Payment Timing		By Tax Type	
	First Year (1)	Second Year (2)	Retroactive (3)	Concurrent (4)	Non - VAT (5)	VAT + Non-VAT (6)
a. Audit - Statitstics (N= 10,272)						
Audit Probability (%)						
Post-Treatment	0.030	0.170	-2.136**	0.162	0.127	0.094
	(0.236)	(0.236)	(1.040)	(0.246)	(0.440)	(0.266)
Pre-Treatment	0.025	-0.041	-1.984**	0.123	0.192	0.111
	(0.115)	(0.069)	(0.806)	(0.121)	(0.388)	(0.169)
Penalty Size (%)						
Post-Treatment	-0.118	-0.233*	1.044	-0.191*	-0.291	-0.179
	(0.115)	(0.122)	(0.742)	(0.112)	(0.180)	(0.112)
Pre-Treatment	-0.001	0.049	0.032	-0.001	-0.376**	-0.136
	(0.088)	(0.034)	(0.470)	(0.093)	(0.168)	(0.086)
b. Audit - Threat Letters (N= 4,048)						
Audit Probability (%)						
Post-Treatment	0.376*	0.378*	0.831	0.335	0.164	0.295*
	(0.210)	(0.220)	(0.944)	(0.216)	(0.185)	(0.170)
Pre-Treatment	-0.342*	-0.214	0.013	-0.289	-0.219	-0.308**
	(0.178)	(0.162)	(0.647)	(0.182)	(0.163)	(0.142)

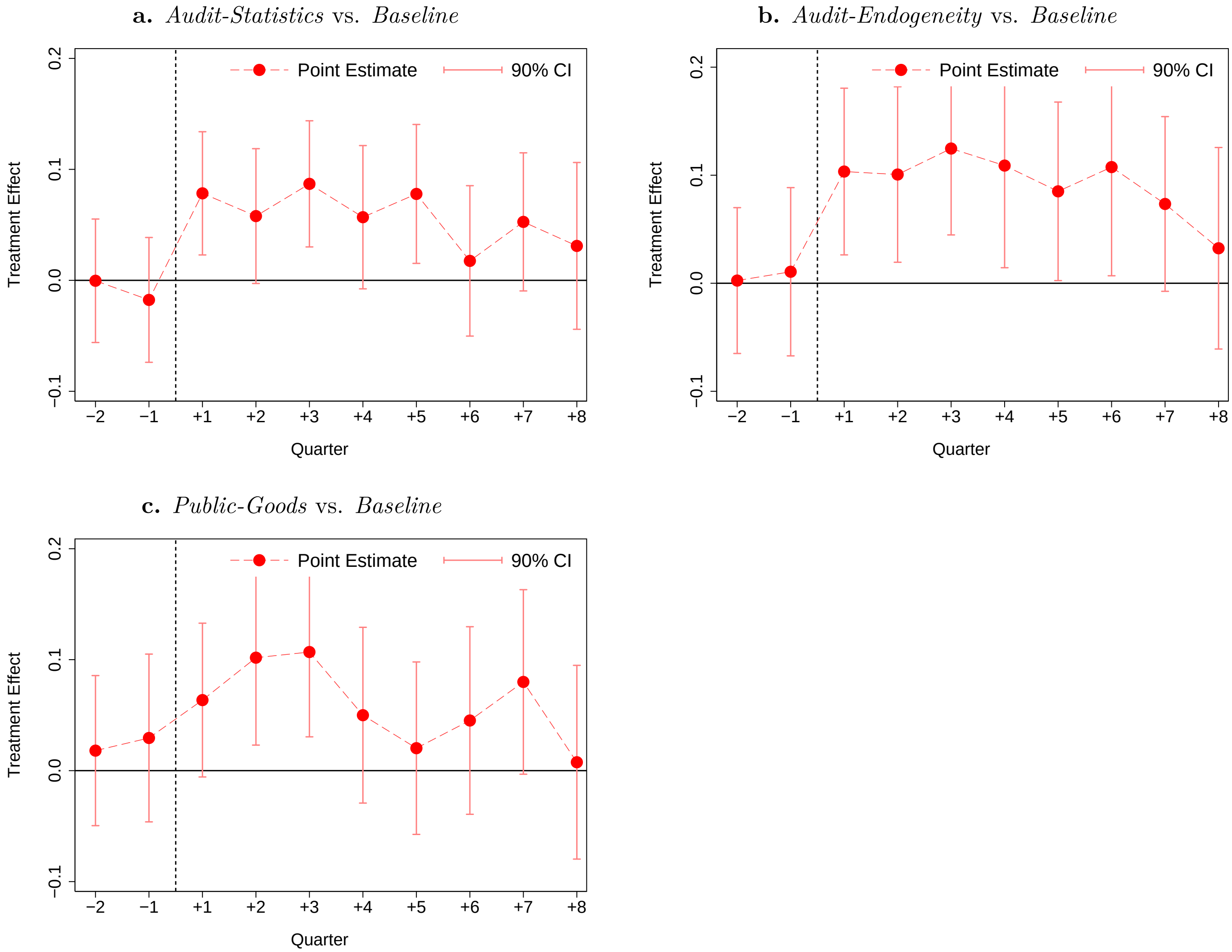
Notes: * significant at the 10% level, ** at the 5% level, *** at the 1% level. Robust standard error.s Panel a. presents the effect of providing different information regarding p and in the *audit-statistics* message. Panel b. compares the two *audit-threat* messages, i.e. the 50% threat of audit vs. the 25% threat of audit. Rows (1) and (3) of panel a. present the effect of informing an additional percentage point of p and θ respectively on the post-treatment VAT payments. The results are based on a Poisson regression, so coefficients can be interpreted directly as semi-elasticities. Regressions are estimated using monthly pre-treatment controls and the stratification variable used to randomize the parameters. Rows (2) and (4) present a falsification test in which we estimate the same regression using pre-treatment information as the dependent variable. All regressions are estimated with a set of monthly controls that correspond to the year before the outcome, i.e. in the post-treatment outcome we include monthly VAT payment controls from September, 2014 to August, 2015 and in the pre-treatment outcome we include the same variables for the September, 2013 - August 2014 period. We also restrict the analysis to firms that effectively received the letter as reported by the postal service. Row (1) in panel b. presents the post-treatment effect of receiving the letter of 50% threat relative to receiving the 25% threat letter. Row (2) of panel b. replicates the estimates for the pre-treatment outcomes. Column (1) shows the first year effect estimates (October 2015 - September 2016) while Column (2) reports the effect of the letter in the second year after the treatment (October 2016 - September 2017). Columns (3) and (4) show the first year effect of treatment on retroactive (3) and concurrent (4) and VAT payments. Columns (5) and (6) report the results by type of tax. Column (5) shows the first year effect of the treatment in other non-VAT tax payments while column (6) reports the effect on the total amount of taxes paid by the firms in the same period.

Figure 1: Distribution of Statistics Shown in Audit-Statistics Letter Type by Size of the Firm



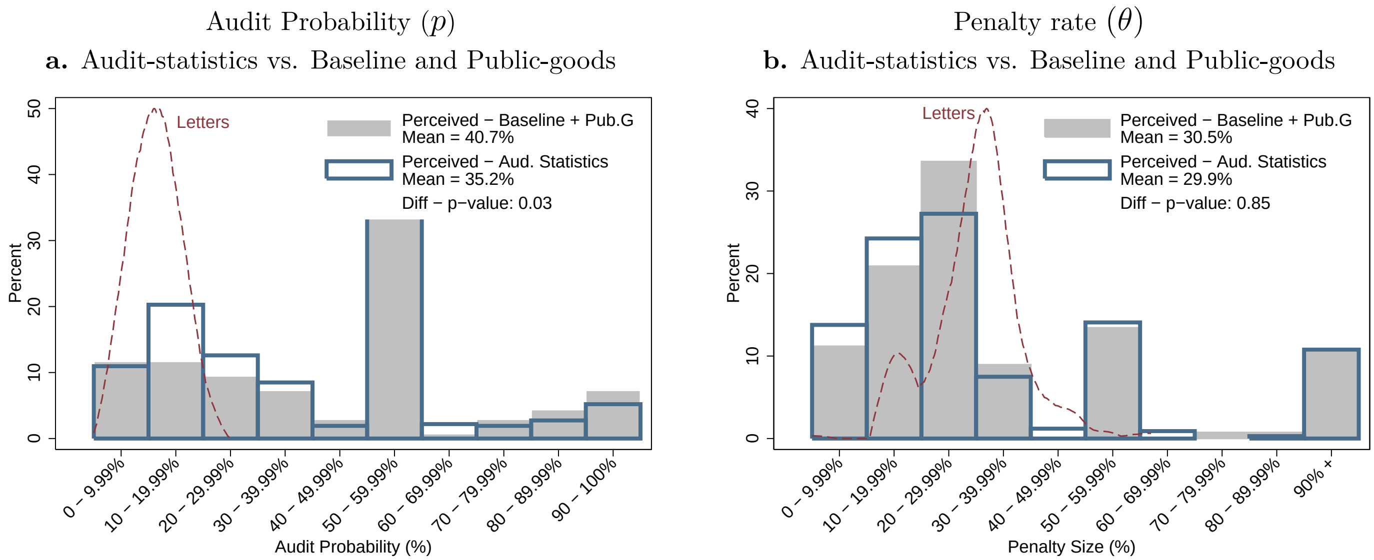
Notes: Notes: N=10,272. Information provided in the audit-statistics letter: probability of being audited (p, in panel (a)) and penalty rate (θ, in panel (b)). Group 1 through 5 correspond to each of the sales quintiles.

Figure 2: Average Effects of *Audit-Statistics*, *Audit-Endogeneity* and *Public-Goods* Messages: By Quarter



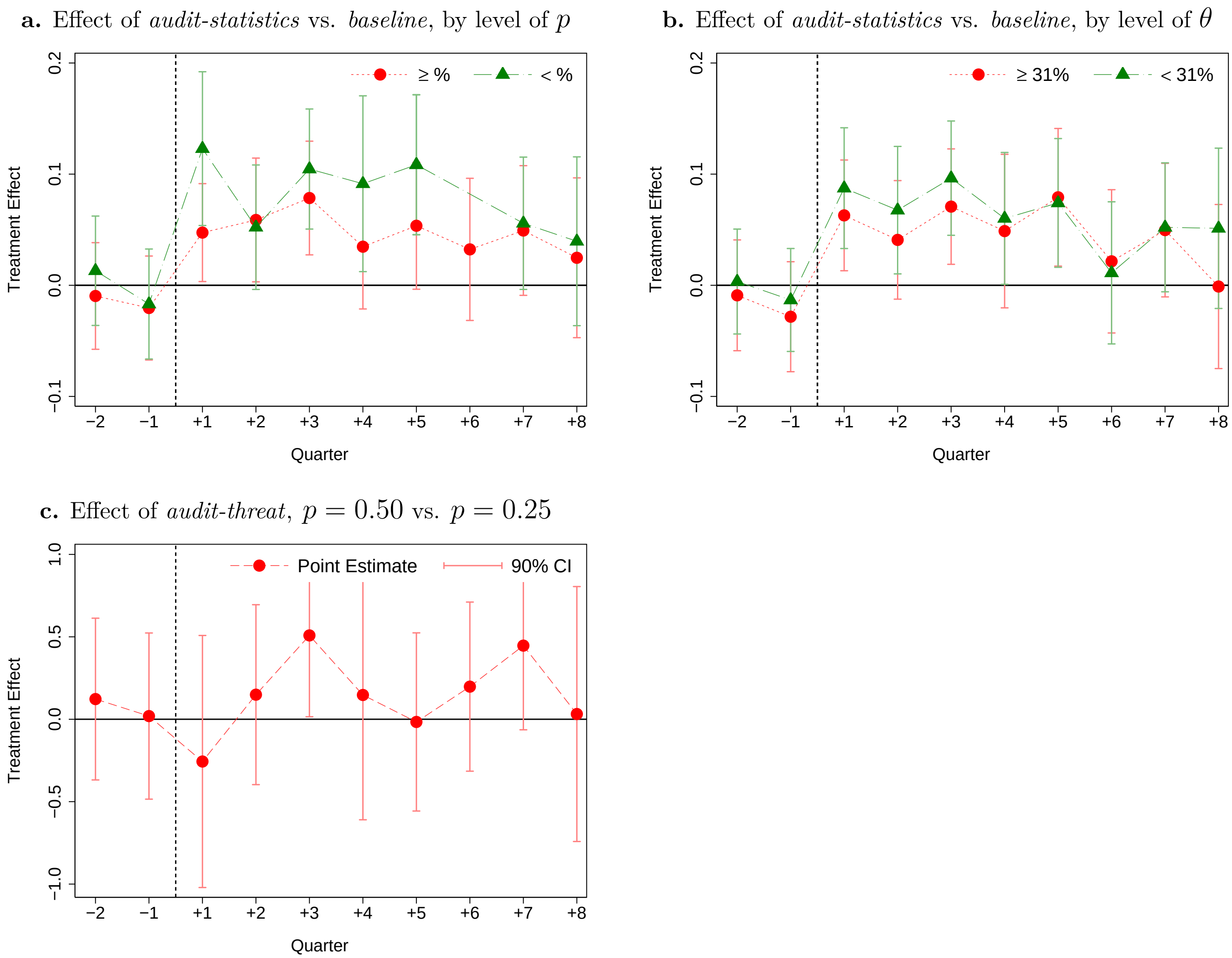
Notes: These figures plot the quarterly effects of each treatment arm when compared to the *baseline letter* . Panel a. (N=12,336) presents the effect of the *audit-statistics* message on total quarterly VAT payments, while panel b. (N=4,103) represents the effect of the *audit-endogeneity* message and panel c. (N=4,081) depicts the effect of the *public-goods* message on the same outcome variable. Each point (red circle) in the plot represents the estimate of the effect of treatment on VAT payments for a specific quarter from two quarters before treatment up to eight quarters after receiving the letter. Regressions do not include monthly pre-treatment controls. The results are based on a Poisson regression, so coefficients can be interpreted directly as semi-elasticities. 90% confidence intervals represented by red lines are computed with heteroskedastic-robust standard errors.

Figure 3: Survey Results: Perceived p and θ, by Treatment Group



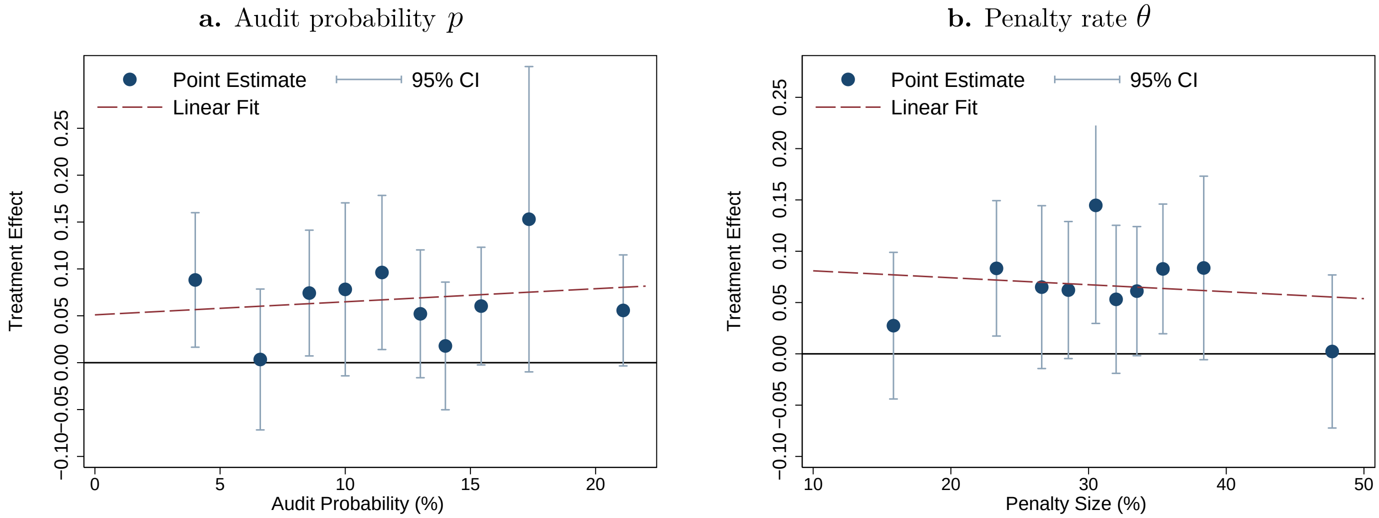
Notes: The histograms are based on survey responses from those self-reported as owners in the post-treatment survey. Perceived - Baseline (N=69) refers to survey respondents who received the *baseline* letter during the experimental stage, while Public-goods (N=68) refers to recipients of the *public-goods* letters. Perceived - Aud. Statistics (N=365) refers to respondents who received *audit-statistics* letters. We also report the p-value of the mean difference between taxpayers that received the *audit-statistics* letter and taxpayers that received the *baseline* letter or the *public-goods* letter. The answers correspond to questions Q2 and Q4 (see full survey questionnaire in Appendix A.7). In panel a. the x-axis represents the probability of being audited; in panel b. it represents the average penalty rate. Perceived Baseline + Pub. G. is the histogram for firm owners who received the *baseline* letter or the *public-goods* letter. Perceived Aud. Statistics refer to the histograms of firms that effectively received *audit-statistics* letters. The red line represents the density function of the information displayed in the *audit-statistics* letters and it is measured in the right y-axis which is hidden for the sake of an easier visualization.

Figure 4: Effects of *Audit-Statistics* and *Audit-Threat* Sub-Treatments: By Deciles of the Signal and by Quarter



Notes: Panel a. (N=10,272) presents the effect of the p reported in the *audit-statistics* message among firms that received the *audit-statistics* letter on the total quarterly VAT payments while panel b. reports the same estimates for θ. Panel c. (N=4,048) depicts the effect of the *audit-threat* message (50% vs 25%). Each dot in panels a. b. and c. represents the estimate of the effect of treatment on VAT payments for a specific quarter from two quarters before treatment up to eight quarters after receiving the letter. Red dots in panels a. and b. represent the effect for *audit-statistics* letter recipients with a reported p or θ above the median. Green dots represent the same effect but for those with reported p and θ below the median. Regressions do not include monthly pre-treatment controls. The results are based on a Poisson regression, so coefficients can be interpreted directly as semi-elasticities. 90% confidence intervals represented by red and green lines, are computed with heteroskedastic-robust standard errors.

Figure 5: Effect of *audit-statistics* vs. *baseline*, by deciles of p and θ



Notes: Panel a. (N=10,272) plots the first-year effect (October 2015 - September 2016) of the *audit-statistics* letter on total VAT payments by decile of p while panel b. reports the same results by decile of θ. Each dot in both panels represents the estimated treatment effect for each decile of the parameter considered. Regressions are estimated using monthly pre-treatment controls and the stratification variable used to randomize the parameters. All effects are depicted jointly with the 95% confidence interval. The results are based on a Poisson regression, so coefficients can be interpreted directly as semi-elasticities. Confidence intervals, are computed with heteroskedastic-robust standard errors.

Online Appendix: For Online Publication Only

A Replication Material: Letters and Survey

This appendix presents samples of the five letter types: *baseline* letter (A.1), *audit-statistics* letter (A.2), *audit-threat* letter (A.3), *audit-endogeneity* letter (A.4) *and public goods* letter (A.5). Additionally, Appendix A.6 presents a sample of the invitation email sent by the IRS to complete the online survey, and Appendix A.7 presents the module from that survey that we included with the questions that are most relevant for our experiment.

A.1 Sample Letter: Baseline Letter

DIRECCIÓN GENERAL IMPOSITIVA

Montevideo, August 20th 2015

Mr../Ms. Taxpayer:

The DGI has the authority to perform inspections (see Art. 68 of the tax code) and routine audits of taxpayers on the basis of crosschecks and assessment of data compiled to detect oversights and inconsistency on tax returns as well as pending tax debts.

The aim of the DGI, and the primary challenge it faces, is to ensure the collection of revenue to sustain life in society. Additionally, its task is to generate a framework of fair and transparent competition where the failure of some to meet their obligations does not have an unfavorable impact on honest taxpayers. In order to meet these goals, inspections are performed in a routine fashion.

Your micro, small, or medium-sized business has been randomly selected to receive this information. *It is solely for your information and its receipt does not require you to present any documentation to the DGI offices.*

We ask you to comply with your tax obligations for the sake of the country we all want, a more and more developed Uruguay with greater and greater social cohesion.

Sincerely,

El Director General de Rentas

Lic. Joaquín Serra

Collection and Controls Division
Internal Revenues Services

A.2 Sample Letter: Audit-Statistics Letter

DGI | DIRECCIÓN GENERAL IMPOSITIVA

Montevideo, August 20th 2015

Mr../Ms. Taxpayer:

The DGI has the authority to perform inspections (see Art. 68 of the tax code) and routine audits of taxpayers on the basis of crosschecks and assessment of data compiled to detect oversights and inconsistency on tax returns as well as pending tax debts.

On the basis of historical information on similar businesses, there is a probability of p% that the tax returns you filed for this year will be audited in at least one of the coming three years. If, pursuant to that auditing, it is determined that tax evasion has occurred, you will be required to pay not only the amount previously unpaid, but also a fee of approximately θ% of that amount.

The aim of the DGI, and the primary challenge it faces, is to ensure the collection of revenue to sustain life in society. Additionally, its task is to generate a framework of fair and transparent competition where the failure of some to meet their obligations does not have an unfavorable impact on honest taxpayers. In order to meet these goals, inspections are performed in a routine fashion.

Your micro, small, or medium-sized business has been randomly selected to receive this information. *It is solely for your information and its receipt does not require you to present any documentation to the DGI offices.*

We ask you to comply with your tax obligations for the sake of the country we all want, a more and more developed Uruguay with greater and greater social cohesion.

Sincerely,

El Director General de Rentas

Lic. Joaquín Serra

Collection and Controls Division
Internal Revenues Services

A.3 Sample Letter: Audit-Threat Letter

DGI | Dirección General Impositiva

Montevideo, August 20th 2015

Mr../Ms. Taxpayer:

The DGI has the authority to perform inspections (see Art. 68 of the tax code) and routine audits of taxpayers on the basis of crosschecks and assessment of data compiled to detect oversights and inconsistency on tax returns as well as pending tax debts.

We would like to inform you that the business you represent is one of a group of firms pre-selected for auditing in 2016. A p% of the firms in that group will then be randomly selected for auditing.

The aim of the DGI, and the primary challenge it faces, is to ensure the collection of revenue to sustain life in society. Additionally, its task is to generate a framework of fair and transparent competition where the failure of some to meet their obligations does not have an unfavorable impact on honest taxpayers. In order to meet these goals, inspections are performed in a routine fashion.

Your micro, small, or medium-sized business has been randomly selected to receive this information. *It is solely for your information and its receipt does not require you to present any documentation to the DGI offices.*

We ask you to comply with your tax obligations for the sake of the country we all want, a more and more developed Uruguay with greater and greater social cohesion.

Sincerely,

El Director General de Rentas
Lic. Joaquín Serra

Collection and Controls Division
Internal Revenues Services

A.4 Sample Letter: Audit-Endogeneity Letter

DGI | DIRECCIÓN GENERAL IMPOSITIVA

Montevideo, August 20th 2015

Mr../Ms. Taxpayer:

The DGI has the authority to perform inspections (see Art. 68 of the tax code) and routine audits of taxpayers on the basis of crosschecks and assessment of data compiled to detect oversights and inconsistency on tax returns as well as pending tax debts.

The DGI uses data on thousands of taxpayers to detect firms that may be evading taxes; most of its audits are aimed at those firms. Evading taxes, then, doubles your chances of being audited.

The aim of the DGI, and the primary challenge it faces, is to ensure the collection of revenue to sustain life in society. Additionally, its task is to generate a framework of fair and transparent competition where the failure of some to meet their obligations does not have an unfavorable impact on honest taxpayers. In order to meet these goals, inspections are performed in a routine fashion.

Your micro, small, or medium-sized business has been randomly selected to receive this information. *It is solely for your information and its receipt does not require you to present any documentation to the DGI offices.*

We ask you to comply with your tax obligations for the sake of the country we all want, a more and more developed Uruguay with greater and greater social cohesion.

Sincerely,

El Director General de Rentas

Lic. Joaquín Serra

Collection and Controls Division
Internal Revenues Services

A.5 Sample Letter: Public-Goods Letter

DGI | Dirección General Impositiva

Montevideo, August 20th 2015

Mr../Ms. Taxpayer:

The DGI has the authority to perform inspections (see Art. 68 of the tax code) and routine audits of taxpayers on the basis of crosschecks and assessment of data compiled to detect oversights and inconsistency on tax returns as well as pending tax debts.

If those who currently evade their tax obligations were evade 10% less, the additional revenue collected would enable all of the following: to supply 42,000 portable computers to school children; to build 4 high schools, 9 elementary schools, and 2 technical schools; to acquire 80 patrol cars and to hire 500 police officers; to add 87,000 hours of medical attention by doctors at public hospitals; to hire 660 teachers; to build 1,000 public housing units (50m2 per unit). There would be resources left over to reduce the fiscal burden. The tax behavior of each of us has direct effects on the lives of us all.

The aim of the DGI, and the primary challenge it faces, is to ensure the collection of revenue to sustain life in society. Additionally, its task is to generate a framework of fair and transparent competition where the failure of some to meet their obligations does not have an unfavorable impact on honest taxpayers. In order to meet these goals, inspections are performed in a routine fashion.

Your micro, small, or medium-sized business has been randomly selected to receive this information. *It is solely for your information and its receipt does not require you to present any documentation to the DGI offices.*

We ask you to comply with your tax obligations for the sake of the country we all want, a more and more developed Uruguay with greater and greater social cohesion.

Sincerely,

El Director General de Rentas

Lic. Joaquín Serra

Collection and Controls Division
Internal Revenues Services

A.6 Sample Letter: Invitation to the Online Survey

Dear Taxpayer:

The DGI's strategic objectives for this period include improving taxpayer services. In 2013, the first Survey on the Costs of Tax Compliance for Small and Medium-Sized Businesses was administered with the support of the Inter-American Center of Tax Administrations (CIAT) and the United Nations (UN). The DGI, in conjunction with a group of academics, has designed a new version of the survey (for more information, visit www.dgi.gub.uy). You can give us your answers on the website where you will find instructions on how to fill out the simple questionnaire; the entire process should take no more than fifteen minutes.

Respond to survey

To address these concerns, a random sample of taxpayers will receive a survey to be answered anonymously.

You are one of the randomly selected taxpayers, which is why you have received this communication. We are grateful for the time and effort you dedicate to assessing this questionnaire and to responding to it as precisely as possible.

Let me assure you that the survey is completely anonymous and the selection of recipients entirely random. The success of this project lies in the precision of your responses. It is on the basis of those responses and the real information they provide that the DGI will be able to hone the design, in the present and in the future, of its strategies to reduce the costs of compliance.

If you have any questions about this questionnaire, please send an e-mail to encuestas@cedlas.org.

We would like to thank you once again for your contribution to this project, which we are sure will benefit all taxpayers.

Sincerely,

Joaquín Serra
Director of the Income Tax Department

PS: If the "Respond to survey" link doesn't open, copy the following address in your browser:https://URL.

A.7 Excerpt from the Online Survey Questionnaire

Introductory Text:

We would like you to respond to a survey about the costs of paying taxes. We hope you have the ten minutes that responding to the questionnaire will require. We are interested in your opinion and hope you will be frank in your responses, which are anonymous and used only for statistical purposes. We would like to thank you for your participation.

Questions Included in Main Module:

Q1) Have you been subject to a DGI audit (inspection or monitoring) at any point in the last three years?

☐Yes.

☐ No.

Q2)In your opinion,what is the probability that the tax returns filed by a company like yours be audited at least once in the next three years (from 0% to 100%)?

☐ %

Q3) How sure are you of your response?

☐ Not at all sure.

☐ A little sure.

☐Somewhat sure.

☐ Very sure.

Q4)Let's imagine that a company like yours is audited and that tax evasion is detected. What, in your opinion, is the penalty (in %) as determined by law that the firm must pay in addition to the originally unpaid amount? For example, a fee of X% means that, for each $100 not paid, the firm would have to pay those original $100 plus $X in fees.

☐ %

Q5) How sure are you of your response?

☐ Not at all sure.

☐ A little sure.

B Additional Results, Specification and Robustness Checks

B.1 Summary Statistics of Tax Payments

Table B.1 describes the tax payments made by firms that received the *baseline* letter during the pre and post treatment period. The pre treatment period covers the year immediately before the treatment (September 2014 - August 2015) and analogously the post treatment period covers the twelve immediate months after the treatment (October 2015 - September 2016). On average, the amount of VAT paid by a firm that received a baseline letter during the pre treatment period was 7,770 USD while the median was around 4,860 USD and the standard deviation 8,070 USD. In the subsequent year, the average amount of VAT paid was 6,470 USD, the median 7,770 and the standard deviation 3,740 USD. This represents a reduction of 16,7% in the average VAT payments when comparing post and pre-treatment periods. Since the group of firms we analyze are mainly small and medium sized firms, this could be explained by a high turnover rate.

Retroactive VAT payments made by the firms show that most of taxpayers do not made this type of payments. Indeed, the 75th percentile of this distribution is 0. The average amount of backward payments made by these firms during the pre treatment period was 400 USD while in during the post treatment period was about 300 USD. This is consistent with the trend observed for the overall VAT payments. This is also observed when we look into other tax payments. On average, firms paid 4,070 USD on behalf of other taxes in the pre treatment period and 3,300 USD in the post treatment period. These amounts correspond to other taxes applied to retail sales of specific goods and corporate taxes among others. The standard deviation in the pre treatment distribution was 8,570 USD in the pre treatment period and decrease to 5,430 USD in the post treatment period.

This descriptive analysis is indicative of the importance of the VAT in the Uruguayan tax structure since it is observed that VAT payments almost doubles the amount of other taxes payments representing more than 60% of the total taxes remitted by the firms.

B.2 Robustness Checks of Baseline Specification

To assess the robustness of the baseline results in Section 5, Table B.2 presents alternative estimates based on different specifications. The first two columns present estimates of the treatment effects based only on the extensive margin of VAT payments: i.e. the outcome is 1 if the firm made at least one payment in the post-treatment period, and 0 otherwise. Column (1) presents results from a linear probability model, while column (2) presents Probit estimates. There is not much variation in the extensive margin: 96% of firms in the sample made positive payments in the post-treatment period. This is a direct byproduct of the selection of the subject pool: we excluded all firms who did not make at least three payments in the 12 months before the treatment assignment. The effects of the three different messages on the extensive margin are close to zero and not statistically significant.

The specifications in columns (3), (4) and (5) of Table B.2 use the amount of VAT payments as the dependent variable. Column (3) corresponds to our baseline Poisson specification. In turn, column (4) presents estimates based on Tobit regressions and (5) presents OLS estimates. The Poisson model has a main advantage in this context: it deals naturally with bunching of payments at exactly zero, while still allowing for the effects to be proportional. The OLS specification, instead, does not deal with the bunching at zero and does not allow for the effects on amounts to be proportional. The Tobit specification is more appropriate than OLS since it takes into account the censored nature of the data at zero, but it does not allow for the effects to be proportional.

The results from columns (3), (4) and (5) of Table B.2 are identical in terms of the signs and statistical significance of the coefficients, indicating that the results are robust to the three alternative specifications. If anything, the effects are statistically more significant when using the OLS and Tobit models. Even though the results from the Poisson, OLS and Tobit models are not directly comparable in terms of magnitudes, they are roughly consistent. For example, the Tobit model suggests an effect of audit statistics of USD 480 (p-value=0.001). Since the average outcome is USD 6,465, this Tobit coefficient amounts to an effect of about 7.4%, which is in the same order of magnitude than the Poisson model, which indicates an effect of *audit-statistics* of 6.3% (p-value=0.013).

Column (6) of Table B.2 reports the results of estimating our model on the final tax liability calculated on data from annual tax returns. The timeframe in this outcome is completely different to the one of the monthly VAT payments used for our baseline results and therefore they do not necesarilly match. However, Table B.2 a. shows that the effect of the *audit-statistic* letter is 6.8%, which is indeed similar in sign and magnitude to our main result. This effect is estimated precisely and statistical significant at a 5% level. The effect of the treatment in the falsification test is indistinguishable from zero. Table B.2 b. shows that the effect of the endogeneity message on the VAT reported in the annual tax return is also similar to the effect on the monthly VAT payments. In this case, the point estimate of the coefficient is 5.6% which is smaller than the baseline results. However, in this case the coefficient is less precise and is not statistically different from zero at conventional levels. Finally, Table B.2 c. depicts how as in most of our specifications, the effect of the public-goods letter on the VAT liability reported in the tax return is null.

B.3 Robustness Checks for Heterogeneity With Respect to Prior Beliefs

As an additional robustness check for heterogeneous results in terms of prior beliefs, in this section we present the effect of treatment by whether $\hat{p}$ is below or above the specific p chosen for the respondent. To estimate this effect we divided the sample in two groups. On the one hand, firms whose prior beliefs about the probability of being audited were low compared to the probability reported in the letter they received. On the other hand, firms whose beliefs were equal or larger than the reported paramenter. Since we divide the groups according to a characteristic of firms that received the *audit-statistics* message both groups are compared to the firms that received the *baseline letter*.

According to the AS (1972) if we increase the taxpayers beliefs about the probability of being audited, their payments should also increase (the opposite is also true if we reduce their perception about p). Therefore, the expected effect of the *audit-statistics* message is positive on firms with $\hat{p} < p$, and a negative effect on firms who have a $\hat{p} \geq p$ .Column (2) in Table B.3 shows that the average effect of the audit-statistics letter on VAT payments in the first year after receiving the letter is 8.3% for taxpayers with relatively low priors. Column (3) reports that the average effect for taxpayers with relatively large priors was 6.9%. Although the effect in firms with relatively low prior is positive and larger in magnitude compare to firms with relatively high prior, the effect on the latter has the opposite sign of the one predicted by the model. Furthermore, differences in magnitude are statistically and economically insignificant. These results, if anything, provide evidence against the AS predictions.

B.4 Alternative Specifications for Audit Probabilities and Fines

We also assess the robustness of the estimated effects of the signals of audit probabilities and penalty rates on post-treatment payments in two different ways.

First, Table B.4 a. replicates the analysis in Table 5 for the five additional specifications that we also used in Table B.2. The results are essentially the same that the ones reported in section 6. The effect of the information provided within firms that received the audit-statistics letter is zero. Both in the extensive and intensive margin, regardless the specification used and even using the tax liability from the annual tax returns as an outcome, all coefficients associated to the treatment variable are statistically insignificant. This is also valid for the effects of the audit-threat message. Conditional on being treate, the information reported in the letter does not affect firms behavior.

Second, Table B.5 presents the results for an alternative specification of the elasticity estimation in Table 5. Instead of estimating the elasticities w.r.t. p and θ separately, we estimate the elasticity w.r.t. $p * \theta$.

$$Y_i = \alpha + \gamma_{p \cdot \theta} \cdot p_i \cdot \theta_i + X_i \delta + \epsilon_i \tag{B.1}$$

As in the model where p and θ were included separately, the elasticity with this alternative specification is statistically and economically insignificant too.

Table B.2 and Table B.5 provides further evidence supporting the fear channel rather than a rational re optimization as a consequence of being exposed to some information about the tax enforcement mechanisms.

B.5 Survey Results: Robustness to the Control Group Definition

To assess the robustness of the survey results we replicate the analysis in section 6.1 but unlike the main results, we now define the control group as only those firms that received the *baseline* letter. Therefore, the *audit-statistics* group is formed by the same 365 firms that in the main results while the control group consists only of 69 firms instead of the 137 that formed the pooled control group.

Figure B.2 replicates Figure 3 a. and 3 b. The shaded gray bars show the distribution of perceptions for individuals who received the *baseline* letter or the *public-goods* letter. The red dashed curve correspond to the distribution of signals sent to the firms in the *audit-statistics* letters. Although slighlty smaller than in the pooled control group, the baseline perceived audit probability of the *baseline* letter group (37.7%) is still substantially larger than the 11.7% that results from our data for the overall sample and this difference is also statistically significant. The results are also consistent with the main results when we look at the perceived penalty size. There are no statistically significant differences between the perceived penalty size by the firms that received the *baseline* letter and our estimates from the overall data.

We also replicate the analysis the average effect of the *audit-statistic* message on the perceived audit probability and penalty size. The average perceived probability of being audited for the control group is 37.7% which is slightly smaller than the one reported by the pooled control group (40.7%). Since this number is closer to the average perceived audit probability for the *audit-statistics* group (35.2%) and we reduced our statistical power (the control group reduced to the half), the mean difference test does not reject that both estimates are equal. However, the results are in the same direction that our main results. If anything, the *audit-statistics* letter reduced the average perceived probability of being audited. The results of the penalty size differences between treated and control groups are also consistent with the main results. The audit-statistics letter did not affect taxpayers perceptions about the average penalty size and in general the prior beliefs of the firms were accurate.

If firms were rational, all these results would imply that firms would have paid less taxes as a consequence of the update in their beliefs. However, this is not what it is observed. Hence, this evidence supports the hypothesis that the fear channel is driving the results rather than a rational re-optimization.

B.6 Beliefs About Audit Endogeneity

As in the case of the *audit-statistics* treatment arm, we conducted a survey of letter recipients in which we included a specific question to assess whether the information provided in the letter had an impact on beliefs about the endogeneity of audits:

> Perceived Audit Endogeneity: "In your opinion, if a firm that evades taxes doubles the amount it is evading, what is the effect on its probability of being audited?" The possible answers

were: It would increase significantly; It would increase slightly; It would not change; It would diminish slightly; It would diminish significantly.

The distribution of responses to this question about the perceived endogeneity of audits is depicted in Figure B.3. The distribution of perceptions in the *baseline* letter suggests that firms were already aware of this endogeneity. Relative to the *baseline* group, there are no statistically significant differences in the distribution of perceptions for the *audit-endogeneity* group (p-value of 0.67). In a scale from 1 to 5, where 1 is "more evasion significantly increases the probability of being audited", and 5 means "more evasion significantly diminish the probability of being audited", the average belief was 1.45 in the *baseline group* and 1.41 for the *audit-endogeneity* group.

Table B.1: Tax Payments: Summary Statistics

	Mean (1)	SD (2)	10th (3)	25th (4)	50th (5)	75th (6)	90th (7)
VAT Amounts							
Post-treatment	6.47	7.77	0.44	1.30	3.74	8.48	16.55
Pre-Treatment	7.77	8.07	0.96	1.99	4.86	10.94	19.73
Retroactive VAT Amounts							
Post-treatment	0.30	1.40	0.00	0.00	0.00	0.00	0.62
Pre-Treatment	0.40	1.76	0.00	0.00	0.00	0.00	0.81
Other Taxes Amounts							
Post-Treatment	3.30	5.43	0.00	0.95	1.81	3.52	7.42
Pre-Treatment	4.07	8.57	0.04	1.43	2.14	4.37	8.72

Notes: The statistics in this table correspond to firms that received the *baseline* letter (N=2,064). The pre-treatment period ranges from September 1, 2014 to August 31, 2015 and the post-treatment period ranges from October 1, 2015 to September 30, 2016. Data on payments comes from administrative data records. VAT amounts corresponds to VAT payments and withholdings. Retroactive VAT amount corresponds to two months or more retroactive VAT payments and withholding, e.g. VAT payments made in March 2016 corresponding to September 2015. Other taxes includes payments for the corporate tax, the wealth tax and other specific taxes to bussiness activity.

Table B.2: Average Effects of *Audit-Statistics*, *Audit-Endogeneity* and *Public-Goods* Messages: Alternative Specifications

	Prob. Making Positive VAT Payments		VAT Payments			VAT Declared
	OLS (1)	Probit (2)	Poisson (3)	OLS (4)	Tobit (5)	Poisson (6)
a. Audit - Statistics (N=10,272 [6,088]) vs Baseline (N=2,064 [1,270])						
Post-Treatment	-0.001	-0.019	0.063**	0.493***	0.480***	0.068**
	(0.004)	(0.066)	(0.025)	(0.140)	(0.147)	(0.030)
Pre-Treatment			-0.008	0.047	0.049	0.011
			(0.021)	(0.128)	(0.128)	(0.031)
b. Audit - Endogeneity (N=2,039 [1,233]) vs Baseline (N=2,064 [1,270])						
Post-Treatment	0.001	-0.006	0.074**	0.591***	0.592***	0.056
	(0.006)	(0.085)	(0.032)	(0.189)	(0.195)	(0.043)
Pre-Treatment			-0.006	0.057	0.060	0.027
			(0.027)	(0.169)	(0.169)	(0.039)
c. Public - Goods (N=2,017 [1,240]) vs. Baseline (N=2,064[1,270])						
Post-Treatment	0.006	0.045	0.043	0.333*	0.357**	-0.014
	(0.006)	(0.087)	(0.030)	(0.171)	(0.177)	(0.040)
Pre-Treatment			-0.004	0.059	0.065	0.056
			(0.026)	(0.162)	(0.162)	(0.042)
Mean of Dep. Var.	0.956			6.465		7.194

Notes: * significant at the 10% level, ** at the 5% level, *** at the 1% level. Robust standard errors. Panel a. compares the *audit-statistics* message with the *baseline* letter while panels b. and c. replicate the analysis for the *audit-endogeneity* and *public goods* messages respectively. In the first row of each panel the dependent variable is the amount of VAT payments (in dollars) in the first year after receiving the letter (October 2015 - September 2016). The second row presents a falsification test in which we estimate the same regression, but using the amount contributed the year before receiving the mailing as the dependent variable (September 2014 - August 2015). The mean of the dependent variable reported in the last row corresponds to the mean of the baseline group. All regressions are estimated with a set of monthly controls that correspond to the year before the outcome, i.e. in the post treatment outcome we include monthly VAT payment controls from September, 2014 to August, 2015 and in pre-treatment outcome we include the same variables for the September, 2013 - August 2014 period. We also restrict the analysis to those firms that effectively received the letter. Columns (1) and (2) show the treatment effect on the extensive margin using two alternative strategies. Column (1) presents the treatment effect on the probability of making at least one VAT payment in the post treatment period using a OLS model. Column (2) replicates the same analysis using a probit model. Columns (3), (4) and (5) present different estimation strategies for the intensive margin, i.e. the total amount of VAT paid. In column (3) we present the results of a Poisson estimation, while column (4) uses an OLS regression and column (5) depicts the Tobit estimation results. Column (6) reports the result of estimating a poisson model on the final VAT liability calculated from annual tax returns. The last row of the table presents the baseline mean for each group of outcomes.

Table B.3: Effects of Audit-Statistics: By Reported Audit Probability

		By $\hat{p}$	
	All	$\hat{p} < p$	$\hat{p} \geq p$
	(1)	(2)	(3)
a. Audit - Statistics vs Baseline			
Post-Treatment	0.063**	0.083***	0.069***
	(0.025)	(0.027)	(0.025)
Pre-Treatment	-0.008	0.003	0.009
	(0.021)	(0.014)	(0.015)
Observations	12,336	6,776	7,624

Notes: * significant at the 10% level, ** at the 5% level, *** at the 1% level. Robust standard errors. Panel a. compares the *audit-statistics* message with the *baseline* letter. In the first row the dependent variable is the amount of VAT payments (in dollars) in the first year after receiving the letter (October 2015 - September 2016). The second row presents a falsification test in which we estimate the same regression, but using the amount contributed the year before receiving the mailing as the dependent variable (September 2014 - August 2015). All regressions are estimated with a set of monthly controls that correspond to the year before the outcome, i.e. in the post treatment outcome we include monthly VAT payment controls from September, 2014 to August, 2015 and in pre-treatment outcome we include the same variables for the September, 2013 - August 2014 period. We also include the stratification variable used to randomize the parameters. The analysis is restricted to those firms that effectively received the letter. Column (1) shows the baseline result, i.e. the first year effect estimates (October 2015 - September 2016). Column (2) reports the effect of the audit-statistics message on firms whose prior beliefs about the probability of being audite were below the p reported in the letter that they received. Column (3) reports the results for firms whose prior beliefs were above the reported p. Because the *baseline letter* was the comparison group in both columns, the sum of the number of observations in each regression does not add up to 12,336, which is the total number of firms considered in the baseline analysis.

Table B.4: Effects of *Audit-Statistics* and *Audit-Threat* Sub-Treatments: Alternative Specifications

	Prob. Making Positive VAT Payments		VAT Payments			VAT Declared
	OLS (1)	Probit (2)	Poisson (3)	OLS (4)	Tobit (5)	Poisson (6)
a. Audit - Statistics (N=10,272 [6,088])						
Audit Probability (%)						
Post-Treatment	0.007	0.376	0.030	-0.441	0.141	-0.202
	(0.040)	(0.619)	(0.236)	(1.617)	(1.728)	(0.244)
Pre-Treatment			0.025	0.183	0.212	-0.056
			(0.115)	(0.943)	(0.945)	(0.163)
Penalty Size (%)						
Post-Treatment	0.002	0.079	-0.118	-0.362	-0.319	-0.114
	(0.021)	(0.316)	(0.115)	(0.856)	(0.896)	(0.134)
Pre-Treatment			-0.001	-0.123	-0.133	0.226***
			(0.088)	(0.785)	(0.785)	(0.087)
b. Audit - Threat Letters (N=4,048 [3,236])						
Post-Treatment	0.009	0.160	0.376*	1.325	1.253	0.450**
	(0.026)	(0.320)	(0.210)	(0.900)	(0.940)	(0.203)
Pre-Treatment			-0.342*	-0.958	-0.954	-0.528*
			(0.178)	(0.695)	(0.699)	(0.315)
Mean of Dep. Var.	0.943			6.465		7.194

Notes: * significant at the 10% level, ** at the 5% level, *** at the 1% level. Robust standard error. Panel a. presents the effect of providing different information regarding p and θ in the *audit-statistics* message. Panel b. compares the two *audit-threat* messages, i.e. the 50% threat of audit vs. the 25% threat of audit. Rows (1) and (3) of panel a. present the effect of informing an additional percentage point of p and θ respectively on post treatment VAT payments in the first year after receiving the letter (October 2015 - September 2016). Rows (2) and (4) present a falsification test in which we estimate the same regression but using VAT payments during the year before receiving the mailing as the dependent variable (September 2014 - August 2015). The last row corresponds to the mean of the baseline group. All regressions are estimated with a set of monthly controls that correspond to the year before the outcome, i.e. in the post treatment outcome we include monthly VAT payment controls from September, 2014 to August, 2015 and in the pre-treatment outcome we include the same variables for the September, 2013 - August 2014 period. We also restrict the analysis to firms that received the letter according to the postal service. Row (1) in panel b. presents the post treatment effect of receiving the letter of 50% threat relative to receive the 25% letter. These coefficients can be interpreted as elasticities. Row (2) of panel b. replicates the estimates for the pre-treatment outcomes. Column (1) presents the treatment effect on the probability of making at least one VAT payment in the post treatment period using a OLS model. Column (2) replicates the same analysis using a probit model. Columns (3), (4) and (5) present different estimation strategies for the intensive margin, i.e. the total amount of VAT paid. In column (3) we present the results of a Poisson estimation, while column (4) uses an OLS regression and column (5) depicts the Tobit estimation results. Column (6) reports the result of estimating a poisson model on the final VAT liability calculated from annual tax returns. The last row of the table presents the baseline mean for each group of outcomes.

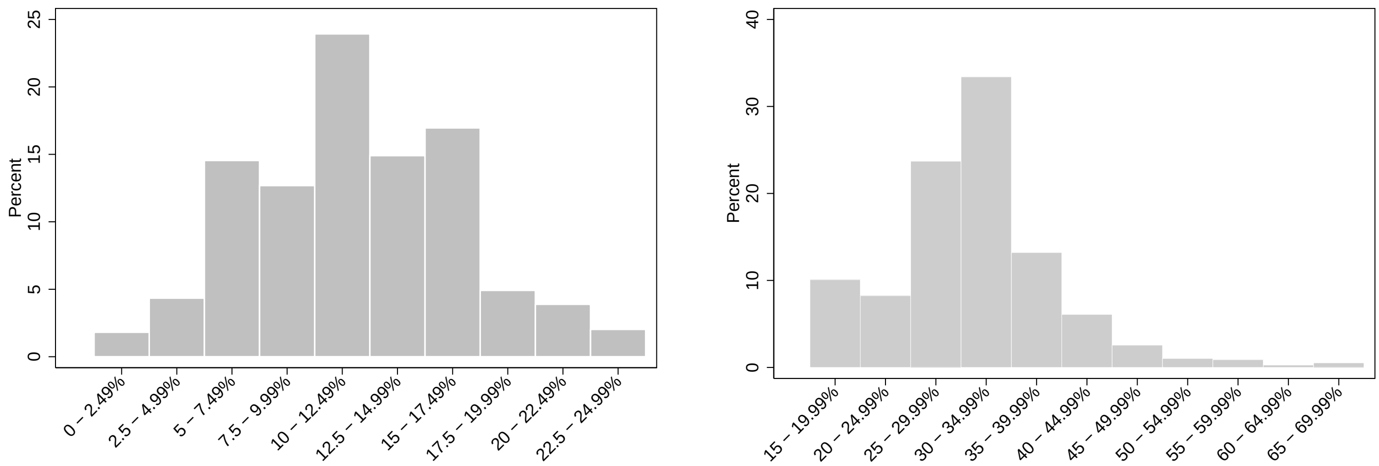
Table B.5: Effects of Audit-Statistics Sub-Treatments: Alternative Specification of $p \cdot \theta$

	By Time Horizon		By Payment Timing		By Tax Type	
	First Year (1)	Second Year (2)	Retroactive (3)	Concurrent (4)	Non - VAT (5)	VAT + Non-VAT (6)
a. Audit - Statistics Letters (N=10,272)						
p*θ (%)						
Post-Treatment	-0.373	-0.537	-1.839	-0.290	-0.902	-0.511
	(0.530)	(0.549)	(3.003)	(0.543)	(0.938)	(0.562)
Pre-Treatment	0.049	0.112	-4.028**	0.249	-1.001	-0.282
	(0.316)	(0.160)	(2.012)	(0.327)	(0.827)	(0.374)

Notes: * significant at the 10% level, ** at the 5% level, *** at the 1% level. Robust standard error. Panel a. presents the effect of providing different information regarding $p^*\theta$ in the *audit-statistics* message. Row (1) of panel a. present the effect of informing an additional percentage point of p and θ respectively on post treatment VAT payments in the first year after receiving the letter (October 2015 - September 2016). Row (2) present a falsification test in which we estimate the same regression but using VAT payments during the year before receiving the mailing as the dependent variable (September 2014 - August 2015). All regressions are estimated with a set of monthly controls that correspond to the year before the outcome, i.e. in the post treatment outcome we include monthly VAT payment controls from September, 2014 to August, 2015 and in the pre-treatment outcome we include the same variables for the September, 2013 - August 2014 period. We also restrict the analysis to firms that effectively received the letter according to the postal service. Column (1) shows the first year effect estimates (October 2015 - September 2016) while Column (2) reports the effect of the letter in the second year after the treatment (October 2016 - September 2017). Columns (3) and (4) show the first year effect of treatment on retroactive (3) and concurrent (4) and VAT payments. Columns (5) and (6) report the results by type of tax. Column (5) shows the first year effect of the treatment in other non-VAT tax payments while column (6) reports the effect on the total amount of taxes paid by the firms in the same period.

Figure B.1: Distribution of Statistics Shown in Audit-Statistics Letter Type

a. Audit Probability (p) **b.** Penalty rate (θ)

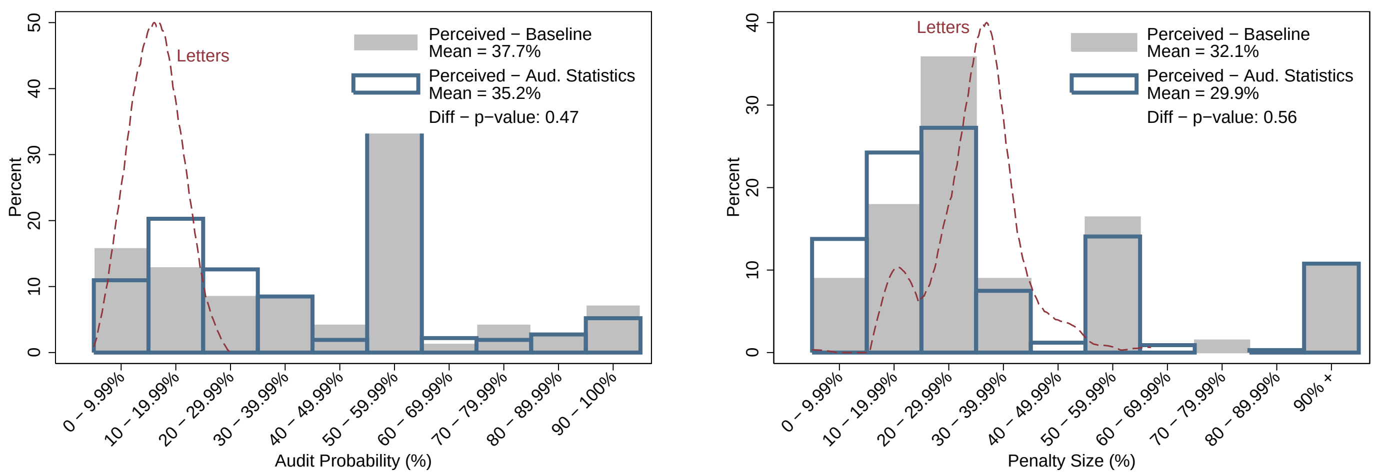


Notes: The data corresponds to the distribution of the information provided in the *audit-statistics* letter (N=10,272). Panel a. presents the distribution of the probability of being audited (p). Panel b represents the distribution of the average penalty rate (θ). p and θ arise from the following procedure: 1) divide the firms in total sales revenues quintiles, 2) randomly draw a sample of 50 similar firms (i.e. from the same quintile) and 3) compute the average p and θ from that sample. This algorithm lead to 950 different combinations of the two parameters.

Figure B.2: Survey Results: Perceived p and θ, by Treatment Group

Audit Probability - p | Penalty rate - θ

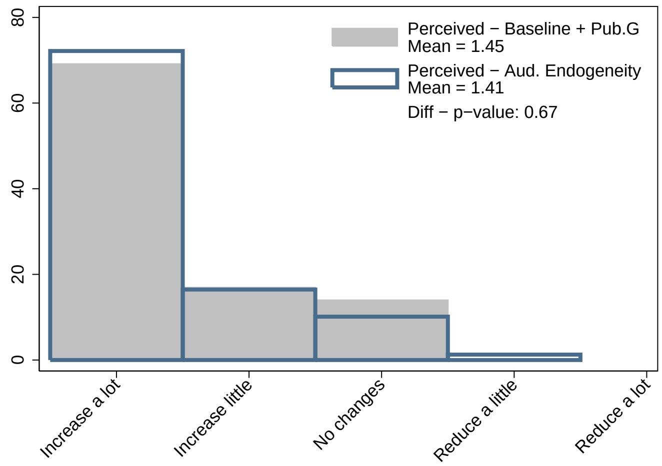
a. Audit-statistics vs. Baseline | **b.** Audit-statistics vs. Baseline



Notes: The histograms are based on survey responses from those self-reported as owners in the post-treatment survey. Perceived - Baseline (N=69) refers to survey respondents who received the *baseline* letter during the experimental stage, while Perceived - Aud. Statistics (N=365) refers to respondents who received *audit-statistics*. We also show the p-value of the mean difference between taxpayers that received the *audit-statistics* letter and taxpayers that received the *baseline* letter. The answers correspond to questions Q2 and Q4 (see full survey questionnaire in Appendix A.7). In panel a. the x-axis represents the probability of being audited; in panel b. it represents the average penalty rate. Perceived Baseline is the histogram for firm owners who received the *baseline* letter. Perceived Aud. Statistics refers to the histograms of firms that effectively received *audit-statistics* letters. The red line represents the density function of the information displayed in the *audit-statistics* letters which is measured in the right y-axis which is hidden for the sake of an easier visualization.

Figure B.3: Perception of Endogeneity of Audits

Audit-endogeneity vs. Baseline and Public-Goods



Notes: Histograms are based on responses to the post treatment survey by respondents self-identified as owners. Perceived - Baseline (N=69) refers to respondents who received the *baseline letter* during the experimental stage, while Public-goods (N=68) refers to recipients of the *public-goods* letters and Perceived Endogeneity (N=79) refers to respondents who received *audit-endogeneity* letters and . These answers correspond to the question Q6 of the survey questionnaire (see the full questionnaire in Appendix A.7). The x axis represents the different categories allowed in the relationship between evasion and probability of being audited question. Perceived Baseline refers to the histogram of the owners of firms that received the *baseline letter.* Perceived Endogeneity refers to the histograms of firms that effectively received *audit-endogeneity* letters.

C Calibration of Allingham and Sandmo (1972)

In this appendix, we compute the predicted elasticities of taxes paid with respect to audit and penalty rates under different calibrations of the *A-S* model. This will provide us with a useful benchmark for the corresponding elasticities estimated with the field experiment.

C.1 Setup

We use an extension of the original *A-S* model. Let Y be the total value-added and let τ be the value added tax rate. Let E be the amount to be under-declared (so $\tau \cdot E$ is the amount evaded). Let p be the probability that the tax declaration for a given year will be audited sometime in the future, and θ as the penalty rate applied over the amount evaded when caught (both of these parameters are defined as in the *audit-statistics* treatment).

The probability of being audited is defined as $p = p_0 + p_1 \frac{E}{Y}$, where $p_1 > 0$ represents the endogeneity of the audit process: i.e., firms that evade more may be more likely to be audited. In the original A-S model the audit probability is exogenous so $p_1 = 0$. In addition to being audited, we assume that firms may be caught evading due to some non-audit technology (e.g., whistle-blowing). As a result, the probability of being caught evading is $p + \epsilon$, where the parameter epsilon represents this non-audit technology.

Each firm has a utility from income given by a Constant Relative Risk Aversion (CRRA) utility function with risk parameter σ. To represent social responsibility, we assume that individuals get some direct utility from paying taxes, equal to the fraction α of the amount paid. This social responsibility parameter α can take values from 0 to 1, where a higher value denotes higher social responsibility. In the original A-S model, $\alpha = 0$.

The optimal evasion is given by maximizing the expected utility:

$$\max_{E \in [0,Y]} \frac{1 - p\left(\frac{E}{Y}\right) - \epsilon}{1 - \sigma} \Big(Y - \alpha\tau(Y - E)\Big)^{1-\sigma} + \frac{p\left(\frac{E}{Y}\right) + \epsilon}{1 - \sigma} \Big(Y - \alpha\tau(Y - E) - (1 + \theta)\tau E\Big)^{1-\sigma}$$

The FOC for the interior solution are:

$$\begin{aligned}\left(1 - p\left(\frac{E}{Y}\right) - \epsilon\right)\Big(Y - \alpha\tau(Y - E)\Big)^{-\sigma}\alpha\tau - \frac{\frac{\partial p\left(\frac{E}{Y}\right)}{\partial E}}{Y(1-\sigma)}\Big(Y - \alpha\tau(Y - E)\Big)^{1-\sigma} - \\ - \left(p\left(\frac{E}{Y}\right) + \epsilon\right)\Big(Y - \alpha\tau(Y - E) - (1 + \theta)\tau E\Big)^{-\sigma}(1 + \theta - \alpha)\tau + \\ + \frac{\frac{\partial p\left(\frac{E}{Y}\right)}{\partial E}}{Y(1-\sigma)}\Big(Y - \alpha\tau(Y - E) - (1 + \theta)\tau E\Big)^{1-\sigma} = 0\end{aligned}$$

$$\left(Y-\alpha\tau(Y-E)\right)^{-\sigma}\left((1-p\left(\frac{E}{Y}\right)-\epsilon)\alpha\tau-\frac{\frac{\partial p\left(\frac{E}{Y}\right)}{\partial E}}{Y(1-\sigma)}(Y-\alpha\tau(Y-E))\right)=$$

$$\left(Y-\alpha\tau(Y-E)-(1+\theta)\tau E\right)^{-\sigma}\left(\left(p\left(\frac{E}{Y}\right)+\epsilon\right)(1+\theta-\alpha)\tau-\frac{\frac{\partial p\left(\frac{E}{Y}\right)}{\partial E}}{Y(1-\sigma)}(Y-\alpha\tau(Y-E)-(1+\theta)\tau E)\right)$$

In the traditional A-S specification, with $\alpha = 1$, $p_1 = 0$ and $\epsilon = 0$, we can obtain a closed analytical form for the elasticities between the VAT payments and p and θ:

$$\frac{\partial log\left(\tau(Y-E)\right)}{\partial p}=-(1-\tau)\frac{-\frac{1}{\sigma}\left(\frac{p\theta}{1-p}\right)^{-\frac{\sigma+1}{\sigma}}\frac{\theta(1+\theta)}{(1-p)^2}}{\left(\theta\left(\frac{p\theta}{1-p}\right)^{-\frac{1}{\sigma}}+1\right)^2\left(\tau-(1-\tau)\left(\frac{\left(\frac{p\theta}{1-p}\right)^{-\frac{1}{\sigma}}-1}{1+\theta\left(\frac{p\theta}{1-p}\right)^{-\frac{1}{\sigma}}}\right)\right)}$$

$$\frac{\partial log\left(\tau(Y-E)\right)}{\partial \theta}=-(1-\tau)\frac{-(1+\theta)\frac{1}{\sigma}\left(\frac{p\theta}{1-p}\right)^{-\frac{\sigma+1}{\sigma}}\frac{p}{1-p}-\left(\frac{p\theta}{1-p}\right)^{-\frac{1}{\sigma}}\left(\left(\frac{p\theta}{1-p}\right)^{-\frac{1}{\sigma}}-1\right)}{\left(\theta\left(\frac{p\theta}{1-p}\right)^{-\frac{1}{\sigma}}+1\right)^2\left(\tau-(1-\tau)\left(\frac{\left(\frac{p\theta}{1-p}\right)^{-\frac{1}{\sigma}}-1}{1+\theta\left(\frac{p\theta}{1-p}\right)^{-\frac{1}{\sigma}}}\right)\right)}$$

When we expand the model with $\alpha < 1$ or $\epsilon > 0$, we can still get closed-form expressions for these two elasticities. However, such elasticities no longer have closed-form solutions when we allow for $p_1 > 0$. In this last case, we use standard numerical methods to compute these elasticities.

C.2 Results

Table C.1 presents the calibration results. Each row corresponds to a different calibration of the *A-S* model. The first seven columns correspond to the parameter values: σ, τ, p_0, p_1, ϵ, θ and α. The last three columns correspond to the predictions of the model under those parameter values. All the parameters are calibrated so that they always predict an evasion rate ($\frac{E}{Y}$) of 26%, which corresponds to the average evasion rate estimated for this country (Gomez-Sabaini and Jimenez (2012)). As a result, all the predictions in the column corresponding to $\frac{E}{Y}$ are the same. We predict the two elasticities discussed in the section above, which correspond to the same elasticities estimated in the experimental analysis. The elasticity $\frac{\partial log(\tau(Y-E))}{\partial p}$ corresponds to the percent change in taxes paid when we increase the audit probability by one percentage point . The elasticity $\frac{\partial log(\tau(Y-E))}{\partial \theta}$ corresponds to the percent change in taxes paid when we increase the penalty rate by one percentage point.

All the calibrations are based on the observed tax rate, $\tau = 0.22$. The calibrations are divided in two groups of four rows each. In the first set of four rows, we assume a CRRA of 4. The model in the first row, sets the audit probability and the penalty rate equal to be exogenous and equal to the observed averages: $p_0 = 0.113$ and $\theta = 0.303$. Given any reasonable CRRA, this model would predict 100% evasion. As a result, we need to extend the model in some way to accommodate the evasion rate of 26% observed in practice. In the first row, we do this by allowing for a non-audit detection rate of $\epsilon = 0.581$. Under this specification, the elasticity with respect to the audit probability is 4.55 and the elasticity with respect to the penalty rate is 3.48. Given that this is perhaps the simple model, and given that its predictions are in the middle range of all the predictions presented in the table, we consider this as our preferred specification.

In the second row, instead of accommodating the evasion rate of 26% by introducing the non-audit detection rate, we use instead the social responsibility model. Assuming a parameter of $\alpha = 0.195$, we can fit the observed evasion rate of 26%. Under this very different approach, the predicted elasticities are somewhat different, but still in the same order of magnitude: the elasticity with respect to the audit probability is 9.422 and the elasticity with respect to the penalty rate is 1.208.

The third row follows a similar specification from the second row, but we augment it by allowing for an endogenous audit probability. We let $p_0 = p_1 = 0.0896$, which accommodates two important features of the audit probabilities: the effective audit probability turns out to be equal to the observed average probability of 11.3%; and, consistent with the content of the *audit-endogeneity* message, if a firm that does not evade taxes ($\frac{E}{Y} = 0$) would double its audit probability if they decided to evade taxes ($\frac{E}{Y} = 1$). This endogeneity would not be nearly enough on its own to fit the observed evasion rate, so we use again the social responsibility parameter to fit the data by setting $\alpha = 0.229$. This specification shows that introducing the endogenous nature of the audit probabilities cuts the elasticities in about half: the elasticity with respect to the audit probability is 4.96 and the one with respect to the penalty rate is 0.66.

The fourth row follows a similar specification from the second row, but we extend it by allowing individuals to have biased perceptions about the audits: $p_0 = 0.367$ and $\theta = 0.31$. These biases would not be nearly enough on its own to fit the observed evasion rate, so we use again the social responsibility parameter to fit the data by setting $\alpha = 0.590$. Again, we obtain elasticities that are in the same order of magnitude than in the other specifications: the elasticity with respect to the audit probability is 4.02 and the one with respect to the penalty rate is 1.64.

The specifications in the second set of four rows are identical to those in the first set of four rows, except that we assume a CRRA of 2 instead of a CRRA of 4. The results indicate that assuming a higher risk aversion provides the more conservative estimates, because the lower risk aversion creates larger elasticities. In sum, the different specifications produce elasticities with respect to the audit probability between 4.02 and 18.857, and elasticities with respect to the penalty rate between 0.66 and 6.60.

Table C.1: VAT Remittances Elasticities Respect to the Main Parameters. Calibrated in the Actual and Perceived Means

Setup							Predictions		
Risk Aversion	Tax Rate	Detection Rate			Penalty	Social Responsibility			
σ	τ	p_0	p_1	ϵ	θ	α	$\frac{E}{Y}$	$\frac{\partial log(\tau(Y-E))}{\partial p}$	$\frac{\partial log(\tau(Y-E))}{\partial \theta}$
4	0.22	0.113	0	0.581	0.303	1	0.26	4.548	3.475
4	0.22	0.113	0	0	0.303	0.195	0.26	9.422	1.208
4	0.22	0.0896	0.0896	0	0.303	0.229	0.26	4.958	0.662
4	0.22	0.367	0	0	0.31	0.590	0.26	4.021	1.642
2	0.22	0.113	0	0.620	0.303	1	0.26	9.854	6.600
2	0.22	0.113	0	0	0.303	0.169	0.26	18.857	2.111
2	0.22	0.0896	0.0896	0	0.303	0.202	0.26	6.283	0.591
2	0.22	0.367	0	0	0.31	0.534	0.26	8.038	2.829

Notes: Each row corresponds to a different calibration of the A-S model. The first seven columns correspond to the parameter values. The last three columns correspond to the predictions of the model under those parameter values. The predicted evasion rate ($\frac{E}{Y}$) is always 26% because all the specifications were calibrated to match the observed average evasion rate of 26%.