Tax Audits as Scarecrows: Evidence from a Large-Scale Field Experiment

Marcelo Bergolo, Rodrigo Ceni, Guillermo Cruces, Matias Giaccobasso y Ricardo Perez-Truglia

Documento de Trabajo Nro. 254
Noviembre, 2019
ISSN 1853-0168
www.cedlas.econo.unlp.edu.ar

Cita sugerida: Bergolo, M., Ceni, R., Cruces, G., Giaccobasso, M. y Perez-Truglia, R.. (2019). Tax Audits as Scarecrows: Evidence from a Large-Scale Field Experiment. Documentos de Trabajo del CEDLAS Nº 254, Noviembre, 2019, CEDLAS-Universidad Nacional de La Plata.

Tax Audits as Scarecrows

Evidence from a Large-Scale Field Experiment

By Marcelo Bergolo, Rodrigo Ceni, Guillermo Cruces,
Matias Giaccobasso, and Ricardo Perez-Truglia*

September 9, 2021

Abstract

The canonical model of Allingham and Sandmo (1972) predicts that firms evade taxes by optimally trading off the costs and benefits of evasion. However, there is no direct evidence that firms react to audits in this way. We conducted a large-scale field experiment in collaboration with Uruguay's tax authority to address this question. We sent letters to 20,440 small and medium-sized firms that collectively paid more than two hundred million U.S. dollars in taxes per year. Our letters provided exogenous yet nondeceptive signals on key inputs for their evasion decisions such as audit probabilities and penalty rates. Using survey data, we measured the effect of these signals on firms' subsequent perceptions of the auditing process. Using administrative data, we measured their effect on actual taxes paid. We find that providing information on audits had a significant effect on tax compliance, but in a manner inconsistent with Allingham and Sandmo (1972). Our findings are consistent with an alternative model of risk-as-feeling, in which messages about audits generate fear and induce probability neglect. According to this model, audits may deter tax evasion in the same way scarecrows scare birds away.

JEL Classification: tax, evasion, audits, penalties, frictions.

Keywords: C93, H26, K34, K42, Z13.

---

*Bergolo: Universidad de la Republica, Uruguay, and IZA (email: mbergolo@iecon.ccee.edu.uy); Ceni: Universidad de la Republica, Uruguay (email: rceni@iecon.ccee.edu.uy); Cruces: CEDLAS-FCE-Universidad Nacional de La Plata, CONICET, and University of Nottingham (email: gcruces@cedlas.org); Giaccobasso: University of California, Los Angeles, and Universidad de la Republica, Uruguay (email: mgiaccobasso@ad.ucla.edu); Perez-Truglia: University of California, Berkeley (email: ricardotruglia@berkeley.edu). We thank Uruguay's national tax administration (Dirección General Impositiva) for their collaboration. We thank Gustavo Gonzalez for his indisensable support of this research. We thank Joel Slemrod for his valuable feedback. We thank comments from participants in seminars at the University of Berkeley, the University of Michigan, the University of California San Diego, Dartmouth University, Universidad Di Tella, Universidad de la Republica, Universidad de Santiago de Chile, Universidad Católica de Chile, UADE, Corporación Andina de Fomento (Buenos Aires), Banco Central del Uruguay, LACEA 2017, 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, the 2018 LAGV Conference, the 2018 IIPF Annual Congress, the 2019 LACEA BRAIN Conference, and the 2019 National Tax Association meeting. This project benefited from funding by CEF, CEDLAS-UNLP and IDRC. The AEA RCT registration number is AEARCTR-0004593.

1 Introduction

Tax audits are a standard tool that most tax administrations have used throughout history. Audits increase tax revenues directly because firms caught evading must pay taxes on the hidden income as well as penalties. Except in the case of large taxpayers, however, these direct revenues are not enough to make audits cost-effective. Audits play a central role in the deterrence paradigm of tax evasion: the threat of being audited in the future–of being caught evading and having to pay penalties–deters firms from evading taxes in the present.

Audits may be useful to fostering tax compliance, but there is no direct evidence on how firms react to them. The Allingham and Sandmo (1972) model (hereafter referred to as *A&S*) is the canonical model of tax evasion in economics. It is an application of Becker (1968), in which selfish individuals choose whether to engage in criminal activities by calculating expected costs and benefits. In *A&S*, firms choose the optimal amount of income to hide from the tax authority so that the marginal benefits (i.e., the lower tax burden) equal the marginal costs (i.e., the penalties they will be required to pay if caught). This intuition is so deeply ingrained in economic thought that most economists take it for granted. Be that as it may, surprisingly little causal evidence exists on whether real firms react to audits in this profit-maximizing fashion (Alm et al., 1992; Luttmer and Singhal, 2014; Slemrod, 2018). In this study, we provide direct tests of the *A&S* model based on a high-stakes, large-scale field experiment.

We study small and medium-sized firms in Uruguay that are subject to the Value Added Tax (VAT). Though that is a context in which taxpayers should care about the threat of being audited, that is not always the case: tax agencies can sometimes use third-party reporting to automatically detect and rectify tax evasion regardless of whether the taxpayer is audited, thus making the audit threat irrelevant. For instance, the U.S. Internal Revenue Service uses their electronic records to compare the wage amount reported by an individual to the amount reported by the individual's employer. This algorithm automatically rectifies the discrepancies in reporting and sends a notification to the taxpayer with the updated tax amount to be paid. Because the evasion will be caught through the third-party reporting regardless of whether the individual is audited, taxpayers should not care about the threat of being audited (Kleven et al., 2011). On the contrary, in our context of the VAT in a developing country, such automatic cross-checking and rectification does not exist. The VAT paper trail, which consists of non-electronic invoices, can only be scrutinized in the event of an audit.[1] Thus, tax authorities must rely heavily on the threat of audits to discourage VAT

[1]While the VAT requires a paper trail, which is a form of third-party reporting, that paper trail is subject to significant limitations, chief among them the fact that there is no simple algorithm that automatically detects tax evasion. Moreover, the paper trail breaks down when it reaches the consumer (Naritomi, 2019).

evasion (Gomez-Sabaini and Jimenez, 2012; Bergman and Nevarez, 2006).

We collaborated with Uruguay's Internal Revenue Service (hereafter referred to as "IRS") to conduct a natural field experiment with a sample of 20,440 small and medium-sized firms subject to the VAT. For our study, the IRS mailed four different types of letters with information on audits to the owners of each of these firms.[2] Some of the information contained in each of these letters was randomized, with the goal of testing predictions of *A&S*. Using IRS administrative records, we measured the effects of the information contained in the letters on the firms' subsequent compliance with the VAT and other tax liabilities. Additionally, we collaborated with the IRS to conduct a post-mailing survey to capture the effect of this information on these firms' subsequent perceptions of audits.

Following the seminal work by Slemrod et al. (2001), the first part of the experimental design measures how informing taxpayers of tax enforcement affects compliance. Firms were randomized into four different letter types: *baseline*, *audit-statistics*, *audit-endogeneity*, and *public-goods*. The *baseline* letter type included brief and generic tax information that the IRS often includes in its communications with firms. The *audit-statistics* letter type was identical to the *baseline* letter, but contained as well information on the probability of being audited and the penalty rate according to tax administration statistics. The hypothesis is that adding the *audit-statistics* message to the *baseline* letter will deter tax evasion, that is, it will increase post-treatment tax payment. We can compare the effects of this *audit-statistics* message with the effects of other types of messages. The *audit-endogeneity* letter type provided information on a different feature of the auditing process. It was identical to the *baseline* letter except for the inclusion of an additional message on how evading taxes increases the probability of being audited. The *public-goods* letter type was designed as a benchmark message that might increase tax compliance without providing information on tax audits. It was identical to the *baseline* letter, except for the inclusion of an additional message describing the social costs of evasion by detailing the set of public goods that could be provided if tax evasion were lower.

We show that, consistent with Slemrod et al. (2001) and the subsequent literature, informing firms of tax enforcement increases compliance. We find that adding the *audit-statistics* message to the *baseline* letter increases tax payments by about 7.0% in the first post-treatment year; the effects continue into the second post-treatment year, but are only half as large and no longer statistically significant. This effect is economically significant: the estimated average VAT evasion rate in Uruguay is 26% (Gomez-Sabaini and Jimenez, 2012).

Finally, firms can also collude to tamper with the paper trail (Pomeranz, 2015).

[2]Throughout this paper, for simplicity's sake, we refer to firms' perceptions and behavior as a shorthand for the perception and behavior of the firms' owners or managers.

While the tax base is not necessarily fully comparable, this figure implies that the 7.0% increase equals a 27% reduction in VAT evasion. The effect of the *audit-statistics* message (increased tax payments of 7.0% in the first year) is similar in magnitude to the effect of the other message on tax audits (7.1%, for *audit-endogeneity*), but larger and more persistent than the effect of the *public-goods* message (5.1% in the first year, but negligible in the second year).

The second and most important part of the experimental design tests the hypothesis that firms react to information about audits as predicted by *A&S*. We provide two tests of *A&S*. The first test exploits survey data on perceptions of audits. If the *audit-statistics* letter had a positive effect on average tax compliance, for this effect to be consistent with *A&S,* it must be true that the letter increased the average perceived probability of being audited or the perceived penalty rate. To test this hypothesis, we designed a survey, which was sent out months after the firms received the *audit-statistics* and *audit-endogeneity* letters, eliciting perceptions of the probability of being audited and the penalty rate.

The second test of *A&S* is based on heterogeneity in the signals induced by the letters. We included exogenous, nondeceptive variation in the information on audit probabilities and penalty rates in the *audit-statistics* letter. To generate this information, we computed the average probabilities and penalty rates using a series of random samples of fifty firms. This sample size was small enough to introduce non-trivial sampling variation in the average probabilities and fines shown to the subjects. Specifically, a given firm could receive a letter saying that the audit probability is 8%, 10%, or 15%, depending on the sample of similar firms chosen for that particular letter. These random variations in probabilities and penalties allow us to test whether firms evade less when they face higher audit probabilities and higher penalty rates, as predicted by *A&S*.

The second part of the results suggests that the effects of the *audit-statistics* letter are not consistent with *A&S*. Based on the survey data, the results for the first test indicate that the *audit-statistics* message reduced the perceived probability of being audited which, according to *A&S*, would in turn reduce compliance. We find, however, that the *audit-statistics* message actually increased average compliance.

The second test shows that, contrary to the *A&S* prediction, signals of audit probability and penalty rates in the *audit-statistics* message had no differential effect on compliance. The estimated elasticity of tax compliance with respect to audit probabilities and penalty rates is close to zero and precisely estimated. Moreover, we compare our experimental estimates to the results from calibrations of *A&S*. We reject the null hypothesis of *A&S* even under conservative assumptions about how much firms learned from the *audit-statistics* message. These results suggest the presence of probability neglect, i.e., that firms react similarly to the threat of being audited regardless of its actual probability or the penalties involved.

As a complement to the *audit-statistics* treatment arm, we designed a separate treatment arm that created exogenous variation in expected audit probabilities in a more direct way. The *audit-threat* letter type was sent to a separate sample of firms that were pre-selected by the IRS for auditing. We randomly divided this set of firms into two groups, one with a 25% probability of being audited and the other with a 50% probability. The *audit-threat* letter informed firms of the exact audit probability that was assigned to them. Consistent with the *audit-statistics* treatment arm, we find probability neglect in the *audit-threat* arm too.

In sum, we find that informing firms of tax audits increased their tax compliance, but the reaction to the information was inconsistent with the optimal reaction predicted by *A&S*. On average, firms reduced, rather than increased, their perceived probability of being audited. Furthermore, reaction was not heightened when firms were faced with a higher probability of being audited or a higher penalty rate. These results suggest that while firms may comply with tax obligations because of the threat of audit, their response is not necessarily optimal as in *A&S*.

Hence, the question arises as to which alternative model best explains the firms' reactions to audits. Models of salience (Chetty et al., 2009) and prospect theory (Kahneman and Tversky, 1979) can explain some, but not all, findings. As highlighted in recent models of firm evasion (Kleven et al., 2016), agency issues within firms could play a role, but they cannot explain our findings either. Our preferred interpretation is based on the model of risk-as-feelings (Loewenstein et al., 2001). The models used for choice under risk are typically cognitive: people make decisions using some type of expectation-based calculus. The risk-as-feelings model proposes that responses to fearsome situations may differ substantially from cognitive evaluations of the same risks. When fear is involved, responses to risks are quick, automatic, and intuitive, and thus neglect the underlying probabilities (Sunstein, 2003; Zeckhauser and Sunstein, 2010). The model of risk-as-feelings can reconcile all of our key findings. Moreover, we present anecdotal and survey evidence indicating that fear of being audited does indeed play a significant role in tax compliance. We also discuss policy implications for increasing tax capacity.

Our study relates to various strands of literature. First, it forms part of a recent but growing body of literature that uses field experiments in partnership with tax authorities to study the decisions 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 Minnesota Department of Revenue with an enforcement message reported higher income in their tax returns than those who received no letter. Similar messages about tax enforcement have been shown to have positive effects on tax compliance in other contexts (for recent reviews, see Pomeranz and Vila-Belda, 2018; Slemrod, 2018; Alm, 2019).[3] One common interpretation in this literature is that taxpayers react to information on tax enforcement tools and, in line with *A&S*, reduce their evasion to re-optimize their behavior. There is no direct evidence in favor of or against this interpretation, however. We hope to contribute by filling this gap in the literature.

This paper is closely related to a group of studies testing *A&S* predictions in a laboratory setting. For example, Alm et al. (1992) conducted a laboratory experiment in which undergraduate students play a tax-evasion game. Subjects can hide income from the experimenter, but some subjects are randomly selected to be audited and must pay a penalty if they are caught evading. The authors show that, in the game, tax compliance increases significantly with audit and penalty rates, but these effects are economically small, indeed smaller than the effects predicted by optimizing behavior in the context of *A&S*. The laboratory experiment setting of Alm et al. (1992) and similar studies have a number of advantages, such as full control over the rules of the game and freedom to select parameters. These laboratory experiments have two main limitations, however. First, the subjects are typically undergraduate students playing the tax game for the first time, with no real-world experience of paying taxes. In contrast, subjects in our field experiment are experienced firm owners who have been registered with the tax agency, that is, paying taxes, for an average of fifteen years. Second, subjects in the laboratory games typically pay less than USD 10 in tax. In contrast, subjects in our field experiment paid USD 11,800 per year in taxes (to get a sense of this magnitude, the Uruguay's GDP per capita in 2015 was around USD 15,000).[4] We contribute to this literature in two ways. We show that *A&S* does not fare substantially better in a natural context with experienced subjects and high stakes. Additionally, we show that audit threats can still be useful to reduce evasion even if taxpayers don't react to audits optimally.

Our findings also contribute to the more general debate about the determinants of tax compliance. The literature wonders why, among smaller firms and self-employed individuals in particular, evasion rates are so low, given the low detection probabilities and penalty rates (Luttmer and Singhal, 2014). One traditional explanation is tax morale: firms and individuals do not evade taxes because they feel morally obliged to comply (Luttmer and Singhal, 2014). Our evidence suggests an alternative explanation: taxpayers overreact to the threat of audits because their tax decisions are emotional. In other words, audits scare taxpayers into declaring their income truthfully just like scarecrows scare birds away. This would explain why, despite low audit probabilities and penalty rates, most taxpayers still

[3] For example, Slemrod et al. (2001); Kleven et al. (2011); Fellner et al. (2013); Pomeranz (2015); Castro and Scartascini (2015); Dwenger et al. (2016); Perez-Truglia and Troiano (2018).

[4] In the twelve months before our experiment, the firms in our sample paid an average of USD 7,770 in VAT and USD 4,070 in other taxes; the GDP per capita in Uruguay was about USD 15,000 in 2015.

report the threat of audit as a major reason why they don't want to evade taxes (United States Internal Revenue Service, 2018).

The paper is organized as follows. Section 2 discusses the experimental design. Section 3 presents the data sources and discusses the implementation of the field experiment. Section 4 presents the results on the average effect of the *audit-statistics* message. Section 5 presents the two tests of *A&S*. Section 6 discusses the interpretation of the findings. The final section concludes.

2 Experimental Design

Our experiment consisted of a mail campaign sent out by Uruguay's IRS with multiple treatment arms and sub-treatments. Rather than comparing firms that received a letter to firms that did not, all of our analyses are based on comparisons between firms that received letters with subtle variations in content. We can thus control for the potential effects on compliance of simply receiving a letter from the tax authority, even if the letter is just a reminder to report taxable income.

The letters consisted of a single sheet of official IRS letterhead with the name of the recipient in the header and the scanned signature of the IRS General Director at the bottom. These letters were folded, sealed in an official IRS envelope, and sent by certified mail to guarantee direct delivery to the recipient and signature upon receipt. Panel (a) of Figure 1 shows the sample sizes for the different treatment arms detailed below.

2.1 *Baseline* Letter

The *baseline* letter contained information on the goals and responsibilities of the tax authority routinely included in IRS communications with firms. It explained that the individual had been randomly selected to receive this information, that the letter was for informational purposes only, and that there was no need to reply or to provide any documentation to the IRS. Figure 2 provides a sample of the *baseline* letter, with the addition of a placeholder box with the word "MESSAGE" written inside.[5] This box was empty in the *baseline* letter but contained a different message (printed in larger print and boldface) in each of the other letter types.

[5]For a full-page sample of the letter without this placeholder, see Appendix A.1. For the corresponding samples of the *audit-statistics*, *audit-threat*, *audit-endogeneity,* and *public-goods* letter types, see Appendices A.2, A.3, A.4, and A.5, respectively.

2.2 ***Audit-Statistics*** Letter

According to the Allingham and Sandmo (1972) model, we expect risk-averse firms to be interested in information on the audit process, because it helps them optimize their tax-payment decisions and potentially increase their bottom line.[6] The information sent should be particularly valuable in a context where information about audits is limited. It is easy to find online data about factors potentially relevant to firms' decision-making, such as prices, inflation, and exchange rates. Information about tax audit probabilities (and, to a lesser extent, actual penalties paid by evading firms) is much harder to come by. Tax authorities seem to prefer to conceal this information.

In the *audit-statistics* letter type, we added to the *baseline* letter the following paragraph on audit probabilities (p) and penalty rates (θ). This letter type was sent to a random sample of firms:[7]

> "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."

We communicated the probability that firms be audited in at least one of the three following years because IRS experts have found that this is the probability that matters to firms as they make decisions. Uruguayan 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 filing be audited is roughly equal to the probability that the firm be audited at least once over the following three years.

In our sample, the average value of p is 11.7%, and 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 others. 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 11.4% and a base θ of 20%.[8]

[6]We assume that firms in our sample are risk-averse–a safe assumption since we deal mainly with small and medium-sized firms. However, *A&S* has also been generalized to settings with risk-neutral agents (Reinganum and Wilde, 1985; Srinivasan, 1973).

[7]To make the information on audit probability and penalty rate clear and salient, we provided all figures as round numbers.

[8]First, there is a 2.1% probability of being audited in any given year, according to the ratio of returns examined for businesses with no income tax credit and with a reported income of between USD 25,000 and 200,000 (Table 9a of IRS, 2014). Each audit covers the previous three to six years, which implies that the probability that the current year's tax filing be audited at some point ranges from 5.88% to 11.42%.

The goal of this treatment arm was to generate exogenous variation in the firms' perceptions of audit probabilities and penalty rates. Because of legal considerations and other constraints, we could not send different firms different sets of information about these factors. We instead induced nondeceptive, exogenous variations in messages that may have an effect on perceptions by exploiting the sampling variation in statistics about audits and penalties. What we did was divide the firms into five groups according to the five quintiles of total VAT payments in the fiscal year before our intervention. For each firm, we then drew a random sample of fifty other firms from the same quintile (i.e., similar firms), for which we computed the averages of p and θ. This randomization strategy generated nine hundred and forty different combinations of p and θ. These estimates of p and θ were unbiased and consistent with the explanation given in a footnote included in the letter. In other words, the information provided to recipients was nondeceptive. The footnote explained how we 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 ranged from 2% to 25%, with an average of about 11.7%. The values of θ ranged from 15% to 68%, with an average of about 30.6%. Figure 3 presents the audit-probability and penalty-rate distribution across five groups by firm size (one in each row) and the distribution of the generated within-group parameters. The vertical line denotes the average audit probability or penalty rate 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 fifty firms, different members of each group received different signals. For example, panel (a.1) of Figure 3 shows that in group 1 (i.e., the first quintile of firms ranked by total VAT payments), the average p for all group members is 8.2%, whereas the histogram depicts the different signals actually sent to firms within the group. These signals cluster around the average of 8.2%, but they range anywhere from 2.5% to 25%.[9]

Second, the IRS usually imposes a basic penalty of θ=20%, although penalties can be higher depending on the situation.

[9] The within-group average p differs across the different groups, increasing monotonically from 8.2% in the bottom quintile to 13.4% in the top quintile. This implies that some of the variation in the values of p and θ included in the letter was non-random. To estimate the causal effects of the signals p and θ, we must isolate the random variation when analyzing the data. In any case, this aspect of the design is not overly important in practice as most of the variation in signals is due to the sampling variation. For example: regressing p on

2.3 *Audit-Threat* Letter

To complement the evidence from the *audit-statistics* sub-treatment, we implemented an alternative randomization of perceptions of audit probabilities with an *audit-threat* letter. We devised a treatment arm that randomly assigned firms to groups with different probabilities of being audited in the following year. The *audit-threat* letter was identical to the *baseline* letter, except for the addition of the following 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."

This *audit-threat* treatment arm was applied to a separate experimental sample, a group of high-risk firms selected by the IRS audit department. The recipients of the *audit-threat* letter cannot, then, be compared to the recipients of the *baseline* letter. Instead, we randomly assigned the firms in this treatment arm to two groups, one with a 25% probability of being audited in the following year (X=25%) and another with a 50% probability (X=50%). These messages were nondeceptive: the IRS audit department committed to conducting audits in the following year according to these probabilities.

2.4 *Audit-Endogeneity* Letter

The *audit-statistics* and *audit-threat* treatment arms conveyed quantitative information about audit probabilities and penalty rates, but we wanted to incorporate into our research design a message about a different aspect of the audit process as well. Most tax agencies, including Uruguay's, consider firm characteristics when deciding which ones to audit. They assign higher audit probabilities to firms with higher evasion risk. As a result, evading taxes typically increases probability of being audited. This factor was incorporated as a special case in *A&S*, in which audit probabilities were determined endogenously. If unsuspecting firms learn about the endogenous nature of their audit probabilities, they should revise their tax-evasion decisions and reduce the amount of tax evaded.[10]

We used this insight from economic theory to devise the *audit-endogeneity* message about the nature of the audit process. We asked our counterparts at the IRS to use their evasion-risk scores to divide a small sample of firms into two groups: those suspected of evading taxes

---

a set of dummy variables for the pre-treatment VAT quintiles results in an $R^2 = 0.135$. Likewise, regressing θ on the same set of dummy variables results in an $R^2 = 0.009$.

[10]Konrad et al. (2016) present suggestive evidence of this mechanism in the context of a laboratory experiment. They find that compliance increases by 80% when taxpayers face a situation where a suspicious attitude toward a tax officer increases audit probability.

and those not suspected of evading taxes. We then computed the difference in audit rates from 2011–2013 between the two groups. We found that the rates were approximately twice as high for the likely-evaders group. On the basis of this information, we created the message in the *audit-endogeneity* letter type, which was identical to the *baseline* letter except for the addition of the following paragraph:

> "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."

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 without directly mentioning audits. We designed a non-pecuniary message based on the suggestions of IRS staff and authorities (i.e., on what information they expected to be most effective at increasing compliance). In the spirit of the model of Cowell and Gordon (1988), this message provided information on the cost of evasion in terms of the provision of public goods.[11] The *public-goods* letter is identical to the *baseline* letter, except for the addition of the following message:[12]

> "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."

2.6 Survey Design

We designed a survey to be conducted with a sample of firm owners from our main subject pool several months after they received the letters. The IRS, with the support of the

[11] This message is also related to the laboratory experiment from Alm et al. (1992), which finds that one of the reasons people decide to pay taxes is appreciation of the public goods provided by tax revenue.

[12] The content of the message was based on estimates from the following governmental agencies: 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).

Inter-American Center of Tax Administrations and the United Nations, had previously administered a survey on the costs of tax compliance to small and medium-sized firms. We collaborated with the tax authority on the design and implementation of a new survey, which included a module tailored to our research design. The survey also included seven additional modules, designed by the IRS, on the costs of tax compliance and other topics.[13] We partnered with local and international universities to increase respondent confidence and to highlight the fact that the survey was part of a scientific study, not of an audit or compliance exercise by the IRS.

To further ensure trustworthy responses, the IRS assured potential respondents that responses would be anonymous and impossible to trace back to specific individuals or firms. To measure the effect of our experiment on these survey responses, we embedded a code in the survey link to identify which treatment arm of the experiment the recipient was assigned to. While these codes did not uniquely identify any firm, they allowed us to link treatment arms and completed questionnaires without compromising anonymity.

In our survey module, we used the following two questions to assess whether the *audit-statistics* message altered recipients' perception of our letters:[14]

> 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 question, we asked how certain the subject felt about his or her response on a 1–4 scale, from "Not sure at all" (1) to "Very sure" (4).[15]

[13] The email with the invitation to participate in the online survey is reproduced in Appendix A.6.

[14] A screenshot of our survey module is found in Appendix A.7.

[15] We also included a question in the survey to measure the subject's awareness of the endogeneity of audit probabilities. You can find a screenshot of this question in Appendix A.7.

3 Data Sources and Implementation of the Field Experiment

3.1 Institutional Context

Uruguay is a middle-income country in South America (the annual GDP per capita was about USD 15,000 in 2015). Our main focus for the study of tax evasion is the VAT, which represents the largest tax liability for firms in Uruguay and also the largest source of tax revenue. At the time of the study, the VAT rate was 22%,[16] and VAT revenues accounted for nearly half of total tax revenues.[17] Uruguay is not atypical in terms of tax evasion. According to estimates from Gomez-Sabaini and Jimenez (2012), evasion of VAT in Uruguay was around 26% in 2008. This is the third-lowest rate in the nine Latin American countries included in the study, and it is comparable to evasion rates in more developed economies. For example, the evasion rate for Italy in 2006 was estimated as 22% (Gomez-Sabaini and Moran, 2014).[18] Uruguay is not an outlier in terms of tax morale either. According to data from the 2010–2013 wave of the World Values Survey, 77.2% of respondents from Uruguay stated that evading taxes is "Never Justifiable," whereas that proportion is, on average, 68.2% for all other Latin American countries (weighted by population) and 70.9% for the United States.

In some contexts, tax authorities do not need to rely on audits to mitigate tax evasion. For example, the U.S. Internal Revenue Service uses their electronic records to compare the wage amount reported by a taxpayer to the amount reported by his or her employer. Their algorithm automatically rectifies the discrepancies in reporting and sends a notification to the taxpayer with the updated tax amount to be paid. Because evasion will be caught through third-party reporting regardless of whether the individual is audited, the probability of being audited should be irrelevant to taxpayers (Kleven et al., 2011). We focused on a context where tax authorities still rely heavily on the threat of audits to discourage evasion. There is some third-party reporting for the VAT, namely the paper trail of invoices for sales and purchases.[19]

[16] A small number of products considered basic necessities had either a 10% rate or were exempt from the tax altogether.

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

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

[19] Firms can 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 difference can be carried over as a credit for future tax years. While the effects of the VAT should, in theory, be similar to those of a retail sales tax, in practice the two types of taxes differ in some significant ways (Slemrod, 2008).

This type of third-party reporting is highly imperfect, however. Most importantly, there is no automatic cross-checking and rectification of VAT payments–that because the paper trail is non-electronic and thus can only be scrutinized by the tax agency in the event of an audit.[20] The VAT paper trail has other limitations documented in the literature, such as breakdown at the consumer end (Pomeranz 2015; Naritomi 2019). The tax agency has access to other enforcement tools, such as tax withholding, but they have limitations as well. As a result, audits are still one of the main ways the tax agency detects tax evasion in the Uruguayan context (Gomez-Sabaini and Jimenez, 2012; Bergman and Nevarez, 2006).

3.2 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. We used a series of criteria to select our main experimental sample from the remaining firms. First, we excluded some firms at the request of the IRS, among them very small or very large firms that were subject to special VAT regimes. We also restricted the experimental sample to firms that had made VAT payments for at least three different months in the previous twelve-month period and to firms with a total value of at least USD 1,000.[21]

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.[22] Moreover, in very large firms the effect of the information could be substantially diluted, as it may not reach the owner or the individuals making decisions about tax compliance. Thus, we excluded from our subject pool firms with a total value exceeding USD 100,000 during the previous twelve months.

These criteria left a subject pool of 20,471 firms for the main experimental sample. All firms were randomly assigned to receive one of the four letter types according to the following distribution: 62.5% were assigned to the main treatment arm (*audit-statistics* letter), and 12.5% were assigned to each of the three remaining letter types (*baseline*, *audit-endogeneity*, and *public-goods*).[23] After removing the 19.9% of letters returned by the postal service, the

[20] The use of standardized electronic invoicing systems may facilitate and automatize the cross-checking of the VAT trail to detect evasion. No such system is in place in Uruguay.

[21] The sample selection was conducted in May 2015, so the twelve-month period spans from April 2014 to March 2015.

[22] In some cases, owners provide the address of an external accountant rather than their address or their firm's address. We removed from the sample firms that were registered with an accountant's mailing address (the IRS keeps records of addresses for all registered accountants).

[23] The randomization of letter types was stratified by the quintiles of the distribution of VAT payments

final distribution of letter types was as follows: 10,272 received *audit-statistics*; 2,064 received *baseline*; 2,039 received *audit-endogeneity*; and 2,017 received *public-goods* letters (total N = 16,392). The 4,597 firms in the secondary sample were assigned to receive the *audit-threat* letter. Half 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).

Table 1 allows us to compare the balance of pre-treatment characteristics for firms assigned to the different letter types. Columns (1) through (4) correspond to firms in the main experimental sample. 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 small and statistically insignificant. Columns (6) through (8) of Table 1 present a similar balance test for the secondary sample used for the *audit-threat* arm. Again, the characteristics are balanced across the two sub-treatments in the *audit-threat* treatment arm.[24]

3.3 Outcomes of Interest

The letters were mailed by Uruguay's postal service on August 21, 2015, and the vast majority were delivered to taxpayers during the month of September. For that reason, we set September 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 is the total amount of VAT liabilities remitted by taxpayers in the twelve months after receiving the letter.[25] To test for the persistence of our treatment effects, a second period of observation, between October 2016 and September 2017 (or up to two years after the intervention), was estalbished. Panel (b) of Figure 1 depicts a timeline of the experiment and the data collection.

The VAT represented 64.4% of the total tax paid by these firms in the fiscal year that preceded our treatment. The corporate income tax represented 25.3% of total tax paid, the wealth tax 6.5%, and the personal income tax withholding only 3.3%. In a context of sole proprietorships, small enterprises, and micro enterprises, the VAT represents the bulk of firms' tax liability, which is why it is our main focus. We did, however, obtain data from the IRS on the other taxes paid by the firms, which allows us to assess whether firms effectively

over the fiscal year before our intervention.

[24]Appendix B.1 provides descriptive statistics for the firms in our subject pool. Moreover, Appendix B.2 shows that the rate of non-delivered letters is mostly balanced across treatments, with only minor and economically insignificant differences in missing delivery status for the *public-goods* and *audit-endogeneity* treatment arms with respect to *audit-statistics* and *baseline*.

[25]This variable includes all VAT payments, that is, direct VAT payments and indirect VAT withholdings.

changed their overall tax compliance or whether they simply substituted the evasion of VAT for evasion of other taxes.

Significantly, the firms in our sample are mostly small. On average, the total amount of VAT paid by the firms that received the baseline letter was about USD 7,700 for the twelve month pre-treatment period; the amount for the corresponding post-treatment period was approximately USD 6,500. That negative trend could be explained by the high share of small firms in our sample, since small firms tend to have high turnover rates. The size of post-treatment VAT payments varied widely, from USD 440 at the 10th percentile to USD 16,550 at the 90th percentile.[26]

We can further break down firms' VAT payments according to timing, observing the date of transfer to the IRS as well as the month for which the payment was imputed. Firms can backdate payments to cover liabilities from previous periods. As firms typically make VAT payments on a monthly basis, they normally cover the current and previous months, which we call concurrent payments. We classified payments covering liabilities incurred two or more months prior as retroactive payments–that is, adjustments for revisions in past liabilities. About 99.7% of firms made one or more concurrent payment in the twelve-month pre-treatment period, whereas only 23.11% of firms made one or more retroactive payment for this same period.[27]

3.4 Survey Implementation

Since the IRS communicates mainly via post, it has mailing addresses for all registered firms. It also keeps records of email addresses for a subset of firms that have used their online services. We emailed invitations to participate in the survey to all firms in the main experimental sample with a valid email address (N=3,867). We wanted to roll out the survey shortly after the reception of the letters but, for reasons beyond our control, we were not able to do so until May 2016, nine months after the intervention.[28] We find that firms invited to the survey were similar in characteristics to the broader set of firms in the main experimental sample.[29]

[26]See Appendix B.3 for detailed descriptive statistics on the distribution of pre- and post-treatment payments for firms that received the *baseline* letter type.

[27]It should be noted that the retroactive payments do not reflect delinquency or outstanding debts to the tax authority. Overall VAT liabilities are computed on a yearly basis, and firms make monthly payments according to their provisional receipts on a pay-as-you-go basis to avoid a large bill at the end of the fiscal year. Thus, retroactive payments reflect changes in past liabilities. For instance, a firm may have "forgotten" to declare a sale in the past and thus need to send a retroactive payment corresponding to the gap between the original and the updated accounting.

[28]While we would have preferred a shorter interval between the experiment and the survey, its design involved several departments of the tax authority, creating delays.

[29]See Appendix B.1.

The main purpose of the survey was to elicit the beliefs of firm owners. We did not include email addresses repeated more than three times in the full sample, since they likely belonged to accounting firms representing multiple small and medium-sized firms. Even after applying this criterion, the IRS records could not ensure that the registered email address belonged to the firms' owners. We thus asked the survey respondent to self-identify as belonging to one of the following five categories: owner, inhouse accountant, external accountant, manager, or other employee. Of the 3,867 recipients invited to participate in the survey, 948 started to answer the survey (a response rate of 24.5%).[30] Of these 948, 68.9% self-identified as an owner and 23.5% as a non-owner; the remaining 7.6% did not provide a response to this question.[31] Our baseline specification excludes respondents who self-identified as non-owners, though the results are similar if we include only those who identified as owners.[32] As per an IRS request, respondents could skip as many questions as they wanted. We find that 6.6% and 8.6% of respondents skipped the audit probability and penalty questions respectively, which is comparable to the average rate (6.1%) at which they skipped other questions in the survey.[33]

4 Results: Average Effect of Messages

4.1 Effect of the *Audit-Statistics* Message

Our first set of hypotheses concern whether providing letters with information on tax enforcement increases tax compliance. We start by describing the effects of our main treatment, the *audit-statistics* message. The literature suggests that a message of this sort will have a positive effect on tax compliance.[34]

Figure 4 summarizes the raw data before conducting any regression analysis. Panel (a) of Figure 4 corresponds to the effect of the *audit-statistics* message. More precisely, the graph shows the percentage difference in VAT paid between the individuals randomly assigned the

[30]For this calculation, we required that respondents had answered at least the first two questions of the survey.

[31]The non-owner responses are distributed as follows: 6.1% self-identified as an internal accountant, 8.3% as an external accountant, 2.7% as a manager, 6.3% as another type of employee.

[32]See Appendix B.4.2.

[33]The skip rate is the probability of providing an answer once a question in the survey has been reached. Appendix B.4 presents a series of robustness tests, and other tests, as well as analyses of differential response rates by treatment group.

[34]As long as the enforcement messages do not affect the taxpayers' true income, the changes in the total amount of VAT paid measure changes in tax evasion. Given the presence of real effects, however, our estimates provide a lower bound for the impact of tax-enforcement information on compliance. Although real effects are possible in our setting, most of the public finance literature provides evidence that real effects are normally zero or small relative to reporting effects (see for example Saez et al., 2012).

audit-statistics letter and the individuals assigned the *baseline* letter.[35] This figure shows the difference for each bimonthly period for all the months for which we have data, including three pre-treatment years (October 2012 to July 2015) and two post-treatment years (October 2015 to September 2017).[36] By construction, period zero is defined as the period during which the letters were being delivered (August–September, 2015); it is highlighted in the figure with the vertical dashed line. Following Pomeranz (2015), we normalize to the average difference during the entire pre-treatment period.

Panel (a) of Figure 4 shows that the *audit-statistics* message had an economically significant effect on VAT payments. Prior to the delivery of the letters, and due to random assignment, we would not expect to see differences between individuals assigned one type of letter or another. In other words, individuals cannot possibly be affected by messages they have not yet received. As expected, the differences in VAT payments between the *audit-statistics* and *baseline* letter recipients hover around zero in the pre-treatment period. Our hypothesis predicts that, after the delivery of the letter, there will be a positive wedge between individuals in the *audit-statistics* and *baseline* letter groups. Indeed, a positive gap in VAT payments between the two groups does arise immediately after the receipt of the letters. In the first two months after the letter delivery, the difference in VAT payments between the *audit-statistics* and *baseline* letter recipients jumps to 10.4% to then hover between 4.0% and 10.5% for the rest of the first post-treatment year. Starting in the second year, the effect grows weaker over time.

While these results suggest that the *audit-statistics* message increased subsequent VAT payments, a more formal framework is required for statistical inference. We observe the outcome variable both before and after the intervention. The resulting information reduces variance in the error term and thus gains statistical power through a difference-in-differences specification that compares treated firms to control firms and the pre-treatment period to the post-treatment period (McKenzie, 2012). We then follow the econometric specification from Pomeranz (2015). Consider the sample of firms assigned either the *baseline* letter or the *audit-statistics* letter. Let i index firms and $t = \{1, 2\}$ denote time, where $t = 1$ corresponds to the twelve months pre-treatment and $t = 2$ to the twelve months post-treatment. Let Y_{it} be the outcome variable by taxpayer i in period t (e.g. $Y_{i,2}$ could be the total VAT payments by firm i in the twelve-month post-treatment period). D_i^1 is a dummy variable that takes value one if i was assigned to receive the *audit-statistics* letter and zero if it was assigned to receive the *baseline* letter. Let $Post_t$ be a dummy variable that takes the value one if $t = 2$

[35] Appendix B.5 discusses the evolution of VAT payments for the treatment and control groups separately.

[36] Since a number of firms are required to pay VAT on a bimonthly basis and there is a strong seasonal pattern, we group the data into bins of two months. In all results, the amounts are top-coded at the 99.99% percentile to avoid contamination by outliers.

(i.e., after the letters were delivered) and zero if $t = 1$. The regression of interest is as follows:

$$Y_{it} = \alpha_0 + \gamma_1 \cdot D_i^1 \cdot Post_t + \alpha_1 \cdot D_i^1 + \alpha_2 \cdot Post_t + \epsilon_{it} \tag{1}$$

The coefficient of interest is γ_1, which measures the differential effect between the *audit-statistics* letter and the *baseline* letter. When the dependent variable is the amount of tax paid, we estimate a log-linear model, also known as a Poisson regression, for two reasons. First and foremost, in the Poisson regression effects are proportional–indeed, the coefficients can be readily interpreted as semi-elasticities.[37] Second, the Poisson regression naturally accounts for the bunching at zero of the dependent variable. Note that the Poisson regression can be used for a continuous non-negative variable; we do not have to rely on additional functional form assumptions such as equidispersion thanks to the quasi-MLE estimator.[38] In any case, we find that the results are robust to alternative regression models (OLS, Tobit, and Probit).[39] Standard errors are always clustered at the firm level.

Table 2 presents the baseline regression results. Panel (a) compares the *audit-statistics* and *baseline* letters. In column (1), the dependent variable corresponds to the effect on VAT paid. The post-treatment coefficient corresponds to the effect during the twelve months after the delivery of the letter (October 2015 to September 2016), as measured by the coefficient γ_1 from equation (1) above. The post-treatment coefficient of *audit-statistics* (in column (1) of panel (a)) is positive (0.070) and highly significant statistically (p-value = 0.001) and economically, suggesting that the *audit-statistics* message increased VAT payments in the twelve months after the intervention by about 7.0%.[40] To better grasp the magnitude of the effects, we can compare them to some basic benchmarks. The estimated average evasion rate for VAT in Uruguay is 26% (Gomez-Sabaini and Jimenez, 2012), and while the tax base is not necessarily comparable, it does provide a benchmark: the 7% increase in VAT payments amounts to a reduction in the evasion rate of 27% ($= \frac{7.0\%}{26\%}$). The effects of our *audit-statistics* treatment are not directly comparable to the effects of the audit message from Pomeranz (2015) because the messages differed in content and each study covers firms from different countries and with different characteristics. Nevertheless, Table 4 from Pomeranz

[37]The Poisson model can be expressed as follows: $\log(Y_X) = \alpha + \beta X + \varepsilon$. The effect of a unit change in X can be re-expressed in log-units of the dependent variable, $\beta = \log(Y_{X=x+1}) - \log(Y_{X=x})$. Provided this coefficient is small enough, it can be approximated as a percent-change effect: $\beta = \log(Y_{X=x+1}) - \log(Y_{X=x}) \approx \frac{Y_{X=x+1} - Y_{X=x}}{Y_{X=x}}$.

[38]For more details, see for example Chapter 19 of Wooldridge (2010).

[39]See Appendix B.6.1 for the results from these robustness checks.

[40]This percent-effect is based on the following approximation: $\beta = \log(Y_{X=x+1}) - \log(Y_{X=x}) \approx \frac{Y_{X=x+1} - Y_{X=x}}{Y_{X=x}}$. For the sake of simplicity, in the rest of the paper we interpret all of the Poisson coefficients using this same approximation. The exact percent-effect can be calculated exactly using the exponential transformation. For example, the 0.070 coefficient corresponds exactly to a 7.25% effect ($= 100 \cdot (e^{0.070} - 1)$).

(2015) indicates that the deterrence letter led to an increase in VAT payments of 7.6%, a figure similar in magnitude and statistically indistinguishable from the 7.0% effect of our *audit-statistics* message. Moreover, our results are qualitatively consistent with a broader literature that finds messages about enforcement to have an effect 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 TV license fees in Austria (Fellner et al., 2013), individual municipal taxes in Argentina (Castro and Scartascini, 2015), an individual church tax in Germany (Dwenger et al., 2016), and tax delinquencies in the United States (Perez-Truglia and Troiano, 2018).

Table 2 presents a number of robustness checks discussed below. First, we present falsification tests in the spirit of event-study analysis. The pre-treatment coefficients from Table 2 are estimated with a specification identical to the one used for the post-treatment coefficients, except for the use of a "placebo" date for the delivery of the letters: i.e., we simulate that the letters were delivered in August and September 2014 and estimate the "effects" on the VAT paid in the subsequent twelve months (i.e., October 2014 to September 2015). Since the letters had not actually been delivered on that date, we would expect the "effect" of the *audit-statistics* message to be close to zero and statistically insignificant. A finding to the contrary would suggest problems with the specification or the random assignment. As expected, the pre-treatment coefficients for the *audit-statistics* message (column (1) in panel (a) of Table 2) is close to zero (0.009), statistically insignificant (p-value=0.658), and as precisely estimated as the corresponding post-treatment coefficient.

Over time, individuals may forget the information conveyed in the letter, or it may become less salient. Beliefs and perceptions may change for other reasons, for instance new events such as audits and information campaigns. To assess the persistence of the effects, column (2) in Table 2 replicates the analysis for the second year after treatment (October 2016 to September 2017). As expected, the effect of the treatment is half as large as it is during the first year, and no longer statistically significant. These estimates are consistent with the pattern of effects by quarter depicted in panel (a) of Figure 4: the effect decreases gradually over time and is about half as large in the second year as in the first. The timing of the effects is also consistent with previous evidence on the effects of tax enforcement messages. The effects of the main intervention in Pomeranz (2015) were also substantially higher in the first twelve months after the intervention, at which point they fell substantially; by the 18th month, they had largely vanished.[41]

Table 2 presents results for complementary outcomes. Firms in Uruguay, as explained in the previous section, make payments not only for their current liabilities, but also for previous

[41] See, for example, Figure 2 in Pomeranz (2015).

periods either because accounts are revised and past mistakes remedied or because invoices not available at the time of the original payment are now imputed. When firms that engage in tax evasion face a heightened threat of audit, we can expect them not only to increase tax payments (i.e., reduce their evasion) in the future, but also to revise their payments for previous periods to reduce or eliminate past evasion. To shed light on this question, columns (3) and (4) in panel (a) of Table 2 split the effects during the first year between retroactive and concurrent payments. The *audit-statistics* message had an economically and statistically significant effect on both retroactive and concurrent payments: the coefficient corresponding to the *audit-statistics* message is 0.383 (p-value=0.006) for retroactive payments and 0.053 (p-value=0.012) for concurrent payments.[42]

We have so far established that firms in the *audit-statistics* treatment arms increased their VAT payments compared to those in the *baseline* letter group. Our analysis focuses on VAT liabilities, which represent the largest fraction of tax payments made by firms in our sample. Our letters referred to taxes in general, however, not VAT or any other tax in particular, which means the effects we reported for VAT may not actually represent a net increase in tax payments: firms may increase their evasion (i.e., reduce their payments) of other tax liablities, thereby crowding out payments or substituting evasion to other taxes. On the other hand, firms may need to declare higher income in order to declare higher VAT, and thus be required to pay more, not less, in non-VAT taxes. The results in columns (5) and (6) of Table 2 shed light on these issues. Column (5) presents the effects on all other taxes paid (mostly the corporate income tax). The effects on the payments of other taxes are as economically and statistically significant as the effects on VAT payments: the *audit-statistics* message had an effect of 8.6% (p-value=0.019) on other tax payments. Column (6) shows that the results are robust if we look at the effect on the sum of VAT and other taxes: the *audit-statistics* message increased this outcome by a statistically significant 7.3% (p-value<0.001).[43]

[42]While the size of the effect for retroactive payments is larger than for concurrent payments, these differences must be taken with a grain of salt because there are large differences in baseline levels between the two outcomes. In the *baseline* letter group, for example, firms paid an average of USD 300 in retroactive payments versus USD 6,160 in concurrent payments in the post-treatment period.

[43]Appendix B.7 presents both a finer analysis that breaks down the effects of "other taxes" into their three components and a series of additional robustness checks, such as alternative estimation methods (Appendix B.6.1), alternative specifications of the dependent variable (Appendix B.6.1), and a heterogeneity analysis based on firm characteristics such as size, age, and sector (Appendix B.8). Overall, we find that the effects are qualitatively and quantitatively similar across the board.

4.2 *Audit-Endogeneity* Message

The first benchmark for the *audit-statistics* message is the *audit-endogeneity* message; the two are similar in that both provide information on tax enforcement through audits. Panel (b) of Figure 4 shows the raw evolution of VAT payments in the *audit-endogeneity* treatment relative to the *baseline* treatment. The results suggest that the *audit-endogeneity* message also induced a significant increase in VAT payments, and that that increase was similar in timing and magnitude to the effects induced by the *audit-statistics* message depicted in panel (a) of Figure 4. For a more formal statistical analysis, the regression results are presented in panel (b) of Table 2.[44] The coefficient in column (1) indicates that the *audit-endogeneity* message increased subsequent VAT payments by 7.1% (p-value of 0.009).[45] This 7.1% is similar in magnitude and statistically indistinguishable (p-value=0.950) from the corresponding 7.0% effect of the *audit-statistics* message reported in panel (a).[46] The same robustness checks were performed on the effects of the *audit-endogeneity* message as on the effects of the *audit-statistics* message: for example, the "effect" on the pre-treatment year is close to zero (-0.005) and statistically insignificant (p-value=0.868), and the effects during the second post-treatment year are about half as large as during the first year.

4.3 *Public-Goods* Message

Panel (c) of Figure 4 shows the effects of the *public-goods* message on VAT payments. The time series data suggest that the *public-goods* message also had a positive effect on tax compliance, but that effect dissipated a lot more quickly than the effects of the *audit-statistics* message. The corresponding regression results are presented in panel (c) of Table 2. The *public-goods* message increased VAT payments in the first post-treatment year by 5.1%, and the effect is statistically significant (p-value=0.043). The effects of the *public-goods* message in the second post-treatment year, however, were close to zero (0.4%) and statistically insignificant (p-value=0.906). The evidence on the effects of moral messages is mixed, and they appear to work in some contexts (Bott et al., 2020; Nathan et al., 2020; Hallsworth et al., 2017) but not others (Blumenthal et al., 2001; Fellner et al., 2013; Castro and Scartascini, 2015; Dwenger et al., 2016; Meiselman, 2018; Perez-Truglia and Troiano, 2018). A closely

[44]The results from panels (b) and (c) in Table 2 are based on a regression specification equivalent to the one from equation (1) above, which was used to obtain the estimates in panel (a).

[45]The estimate of the effect of the *audit-endogeneity* message is somewhat less precise than the effect of the *audit-statistics* message, but that is as expected due to the difference in sample sizes.

[46]This is an equality test between two coefficients based on the same data but different regressions. To allow for a nonzero covariance between these two coefficients, we estimate a system of seemingly unrelated regressions. In the remainder of the paper, when comparing coefficients from the same data but different regressions, we will use this method.

related study, Pomeranz (2015), included a message of moral suasion that had a positive effect on subsequent VAT payments, although that effect was statistically insignificant and not as large as the effect of the deterrence message. Our findings on moral suasion fall closest to the findings of the experiment with Norwegian taxpayers reported in Bott et al. (2020): they find that the message of moral suasion increased tax compliance in the short term, but the effects dissipated completely the following year.

5 Tests of *A&S*

The results presented in the previous section are broadly consistent with the evidence in the literature, namely, providing information on audits significantly increases tax compliance. In this section, we present additional evidence to establish whether the effects of the *audit-statistics* treatment are driven by the *A&S* mechanism.

5.1 First Test of *A&S*: Effects on Perceptions

According to *A&S*, the *audit-statistics* message would have a positive effect on tax compliance if it increased the perceived probability of being audited, the perceived value of the evasion fine, or both. We explore this hypothesis by utilizing data from our post-treatment survey, with 365 firms in the *audit-statistics* group and 137 in what we refer to as the pooled control group, that is, individuals who did not receive information on audits.[47]

Panels (a) and (b) of Figure 5 depict the distributions of perceptions of audit probabilities and penalty rates as elicited from the survey. The shallow bars with solid borders correspond to the perceptions of firms that received the *audit-statistics* message. The shaded gray bars depict the distribution of perceptions of firms in the pooled control group. The red dashed curve, in turn, corresponds to the distribution of signals sent to firms in the *audit-statistics* letters. A comparison of the shaded bars and the red curve from panel (a) of Figure 5 suggests that, on average, respondents in the control group substantially overestimated the probability of being audited. While our administrative data on audits indicate a probability of being audited of about 11.7%, the mean perception for the control group is 40.7% (p-value<0.001 for the difference). This finding of an overestimation of audit probabilities is consistent with prior survey evidence (Harris and Associates 1988; Erard and Feinstein 1994; Scholz and

[47]The size of the survey sample was substantially smaller than the size of our experimental sample. To increase the statistical power of our test, we pooled subjects from the *baseline* and the *public-goods* groups to make up the control group, since both received messages with no specific information on audit probabilities or fines. Appendix B.4.2 shows that the results are similar, but less precisely estimated, when the control group includes only recipients of the *baseline* letter.

Pinney 1995).[48] A comparison of the shaded bars and the red curve from panel (b) of Figure 5, meanwhile, suggests that the average belief about the penalty rates was unbiased: the actual average penalty computed from administrative data for the experimental sample is 30.6%, while the mean perceived penalty for in the control group is 30.5%.

The positive bias in the perceived audit probability could be explained by the availability heuristic bias (Kahneman and Tversky, 1974), according to which individuals judge the probability of an event on the basis of how easily they recall instances of it happening. Even though audits are rare, their likely salience in memory or frequent discussion by colleagues and the media may induce firms to perceive them as 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 (probabilities of dying in a terrorist attack or in an airplane accident (Lichtenstein et al., 1978; Kahneman et al., 1982)).

The survey data indicate that the effects of the *audit-statistics* message are inconsistent with the *A&S* predictions. According to *A&S,* if taxpayers overestimate the audit probabilities on average, the *audit-statistics* message would have *reduced* average tax compliance. The results presented in the previous section show that, on the contrary, the *audit-statistics* message *increased* average compliance.[49] To bolster this argument, we show that the *audit-statistics* letter did indeed have a negative effect on perceived audit probability.[50] The shallow bars with solid borders in panel (a) of Figure 5 show the distribution of perceptions for respondents who received the *audit-statistics* letter. An inspection of panel (a) of Figure 5 indicates that recipients of the *audit-statistics* letter reported a lower perceived probability of being audited, from an average of 40.7% in the pooled control group to an average of 35.2% in the *audit-statistics* group (p-value of the difference 0.03).[51] The mechanisms behind the *audit-statistics* message are relevant to the interpretation of the *audit-endogeneity* message: subjects may have learned that the audits are endogenous through the *audit-endogeneity*

[48]The prior survey evidence was based on responses from wage earners, however, for whom misperception of audit probabilities is largely inconsequential due to widespread third-party reporting (Kleven et al., 2011). On the contrary, the financial stakes of misperceiving audit probabilities can be substantial in our context.

[49]One caveat of the test presented in this section, and for the test presented in the section below, is that we are estimating the effects on the average firm. The fact that the average firm does not behave as *A&S* predicts does not imply that none of the firms do. It is possible, for instance, that some firms altered their perceived probability upward because of the information contained in the letter and increased their tax payments as a consequence. The fact that the average effects are so far from the *A&S* prediction suggests, however, that the firms behaving as *A&S* predicted must have been a minority at most.

[50]One caveat here: a reduction in the self-reported probability of being audited could be caused by an increase in tax compliance due to the endogenous nature of p with respect to tax evasion.

[51]Panel (b) of Figure 5 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 pooled control group to an average of 29.9% for the *audit-statistics* group. The difference is statistically insignificant (p-value of 0.82). For more details, see Appendix B.4.3.

message and re-optimized their tax evasion accordingly; or they could have had a knee-jerk reaction to any information about audits, even if they were already aware of the endogeneity component. Consistent with the evidence on the *audit-statistics* message, we find that the effect of the *audit-endogeneity* message was probably not due to a change in recipients' beliefs: recipients were already aware of the endogeneity.

5.1.1 Concerns with Survey Data

This first test relies on survey data, and as such it faces some challenges common to this type of data. In this section, we discuss and address some of those challenges.

One potential concern is that the responses on audit probabilities and penalty rates mostly reflect measurement errors because the questions were not incentivized. There are several pieces of evidence suggesting otherwise. First, the fact that the survey beliefs changed depending on the information provided in the letters suggests that these responses contained some truthful information. Second, while there is a large positive bias in the perception of audit probabilities, the average perception of the penalty rate (30.5%) is extremely close to the actual probability computed from administrative data (30.6%). The fact that belief and reality line up to such an extent suggests that individuals responded honestly and thoughtfully. Furthermore, individuals were better informed about penalty rates than about audit probabilities, which is also consistent with the fact that there is more readily available information about penalty rates: audits are relatively rare events, and their probabilities are not advertised, whereas evasion penalties are more widely broadcast by the tax agency.

Another potential issue is that respondents were aware that they were misinformed and would never have acted on their biased beliefs. Our survey data provide evidence to the contrary. Even though their estimates were substantially off, survey participants reported being confident of their responses. For example, only 16.2% of those in the control group reported being "Not at all sure" about their perceived probability of audit (on a four-point scale, ranging from "Not at all sure" to "Very sure"); a similar share (18.1%) reported being "Not sure at all" about their guess of the penalty rate. Even for the subgroup of individuals in the control group who reported being "Very sure" of the audit probability, their average belief was, if anything, slightly more biased: they reported a perceived audit probability of 45.7%, which is still substantially higher than the actual probability of 11.7%.

Another concern is that our subjects may have been confused about the questions; perhaps they did not understand the definition of an "audit." Of the 137 responses from the pooled control group, 10.2% of firms reported having been audited in the past three years. This share (10.2%) is close to the actual share of firms that were audited (11.7%), thus suggesting that respondents correctly understood the definition of an audit. Moreover, if firms use their own audit history to form their beliefs about audit probabilities, the ones that had been audited recently should report a higher perceived probability. Indeed, we found that to be the case: firms that had recently been audited reported a substantially higher average perceived probability of being audited in the future (63.9%) than firms that had not been audited recently (38.1% – p-value of the difference<0.001). Subjects could, conceivably, have cognitive limitations when responding to questions about percentages and probabilities. That would be a minor concern in our subject pool, since it is comprised of business owners likely familiar with fractions and probabilities. While no verifying administrative data is available, the anecdotal evidence indicates that our sample is a highly educated subgroup of the population. After all, these business owners need, at the very least, some rudimentary arithmetic skills and understanding of percentages to compute the VAT and other tax liabilities.

Another potential concern is that respondents report a probability of 50% as a way of expressing uncertainty (Bruine de Bruin et al., 2002; Bruine de Bruin and Carman, 2012). Responses of exactly 50% are somewhat common in our data: among individuals in the pooled control group, 41.61% of responses about perceived audit probability and 13.5% about penalty rate are equal to exactly 50%. Our data indicate, however, that most of these responses of exactly 50% are not a product of uncertainty: individuals who provided an answer of 50% are somewhat, but not dramatically, less certain of their responses than individuals who provided answers different from 50%.[52] Moreover, even if we ignored the 50% responses, the main result would still be robust: individuals in the control group still substantially overestimate the probability of being audited (average perception of 34.1%, compared to the actual probability of 11.7%).[53]

As an additional validation of the survey data, we can measure the effect of the signals on p and θ from the *audit-statistics* sub-treatments, though this exercise is limited due to the small sample size (we only have 365 survey responses in the *audit-statistics* group). With that caveat in mind, the survey data suggest that a percentage point increase in the signal on audit probability in the letter increased the perceived audit probability nine months later by 0.397 (SE 0.288) percentage points.[54] While imprecisely estimated and thus statistically insignificant at conventional levels, the magnitude of this belief updating is consistent with the findings from other information-provision experiments (see Section 5.2.1 below for a more detailed discussion). The true effects of the signals in the letter on beliefs were probably

[52] In the pooled control group, the average certainty of perceived audit probability is 2.03 (in the 1-4 scale from "Not at all sure" (1) to "Very sure" (4)) for individuals who responded with a value of exactly 50%, and 2.37 for individuals who responded with a different value (p-value of difference<0.001). For responses about perceived penalty rate, the average certainty is 2.11 for individuals who responded with 50%, and 2.51 for those who responded with another value (p-value of difference=0.001).

[53] For more details, see Appendix B.4.2.

[54] For more details, see Appendix B.4.4.

greater than the above estimates suggest, due to different sources of non-compliance.[55]

We provide an alternative to this test that does not rely on survey data.[56] We assume that, due to the paucity of information publicly available on the auditing process, firms form prior beliefs on audit probabilities based on their own exposure to audits. Take, for instance, two firms that have been paying taxes for ten years where, by chance, one of them has been audited at some point and the other has not. The firm that has been audited in the past will have a higher perceived probability of being audited in the future. The results from this alternative test also provide evidence against the *A&S* mechanism.

5.2 Second Test of *A&S*: Heterogeneity with Respect to Signals

The second test is based on the differential effects of the values of the signals in the letters. According to *A&S*, the effects of the *audit-statistics* message should increase in the signals on audit probability (p) and penalty rate (θ), a hypothesis we tested directly with the random variation we introduced in the p and θ in our *audit-statistics* letter. We first present our estimates of these elasticities, and then compare them with the values obtained from calibrations of the *A&S* model.

5.2.1 Elasticities with respect to p and θ

For a less parametric look at the data, Figure 6 estimates the effect of the *audit-statistics* message on VAT payments, but broken down by decile of the signals in the letter.[57] Panel (a) of Figure 6 presents the effect of the *audit-statistics* message by decile of the signal on p shown in the letter. In the *A&S* framework, we would expect very low signals of p to reduce tax compliance (since they most likely reduced firms' perceived probability of audits); the effect would become larger, and turn positive at some point, as the signal on p is increased. The coefficients plotted in panel (a) of Figure 6, however, indicate that the effect of the *audit-statistics* letter is not related to the value of p included in the letter. The coefficients are

[55]While we are confident that our certified letters reached the firms' owners, we cannot be as confident that the owner was the one who received the email invitation to complete the survey. And while we wanted to conduct the survey shortly after the letter campaign, we were not able to roll it out until nine months after the intervention for reasons beyond our control. In information-provision experiments of this type, the effect of information on beliefs tends to decay substantially in a matter of a few months–recipients may forget the information provided in the letters, or acquire additional information in the meantime. For example, Cavallo et al. (2017) show that the effect of information on beliefs decays by about half in a matter of just three months, and similar findings are reported by Bottan and Perez-Truglia (2017) and Fuster et al. (2018). All these factors will lead to an underestimation of the effects of the letter on beliefs.

[56]For more details, see Appendix B.9.

[57]These results are based on the same specification used for Table 2, except for the addition of dummies for the quintiles of pre-treatment VAT payments as additional controls. On that basis, we drew the sample to calculate p_i and θ_i.

similar in magnitude for the whole range of values from $p = 2\%$ all the way up to $p = 25\%$. Moreover, the resulting linear relationship (shown as a dashed red line) has a slope that is close to zero and statistically insignificant. Panel (b) of Figure 6 provides a similar analysis for the heterogeneity by penalty rates (θ) in the letter. According to *A&S*, we would expect a positive relationship between the effect of the *audit-statistics* letter and the value of θ in the letter. Panel (b) of Figure 6 shows evidence to the contrary: the coefficients are similar for the whole range of values from $\theta = 15\%$ to $\theta = 68\%$, and the slope is close to zero and statistically insignificant.

With a more parametric approach, we can quantify the effects of the *audit-statistics* sub-treatments so that they can be contrasted with the quantitative predictions of *A&S*. For the *audit-statistics* treatment arm, we use the following model:

$$\begin{aligned} Y_{it} = \alpha_0 + \gamma_p \cdot p_i \cdot Post_t + \gamma_\theta \cdot \theta_i \cdot Post_t + \alpha_1 \cdot p_i + \alpha_2 \cdot \theta_i + \alpha_3 \cdot Post_t + \\ + \sum_{g=2}^{5} \alpha_{4,g} \cdot I_{\{i \in g\}} + \sum_{g=2}^{5} \alpha_{5,g} \cdot I_{\{i \in g\}} \cdot Post_t + \epsilon_{it} \end{aligned} \tag{2}$$

where $p_i \in (0, 1)$ is the signal on audit probability included in the letter sent to firm i, and $\theta_i \in (0, 1)$ is the signal on penalty rate included in the letter sent to the same firm. The $I_{\{i \in g\}}$ variables correspond to a set of dummies for the quintiles of pre-treatment VAT payments, which are the groups from which we drew the sample of "similar firms" to calculate p_i and θ_i. Including these controls ensures that we only exploit the exogenous variation in p_i and θ_i, that is, the heterogeneity due to sampling variation. $Post_t$ is a dummy variable that takes the value one if the observation corresponds to the post-experiment period, and zero otherwise. Since we are using a Poisson regression model, γ_p and γ_θ can be directly interpreted as elasticities. For instance, an estimate of $\gamma_p = 1$ would imply that a one percentage-point increase in the audit probability conveyed in the letters increased VAT payments by 1%. *A&S* predicts that $\gamma_p > 0$ and $\gamma_\theta > 0$: i.e., that firms' tax payments increase as their perceived probability of audit and rate of evasion penalty increases.

Panel (a) in Table 3 presents the results from the econometric model of equation (2). Column (1) of Table 3 presents estimates of the elasticities of VAT payments with respect to the values of p and θ conveyed in the *audit-statistics* sub-treatments. The elasticity with respect to the audit probability in the first post-treatment year is -0.063 (p-value=0.796). This means that increasing p by one percentage point would decrease VAT payments by a mere 0.063%. The elasticity with respect to the penalty rate is -0.033 (p-value=0.782), which implies that increasing θ by one percentage point would decrease VAT payments by 0.033%. The estimates are close to zero, statistically insignificant at standard levels, and precisely estimated. The degree of precision means that we can rule out even moderate elasticities: the 95% confidence interval for the audit probability excludes elasticities above 0.411, and the 95% confidence interval for the penalty rate excludes elasticities above 0.198.

Significantly, the pre-treatment falsification test does not yield any statistically significant effects, and the results are similar for the other specifications, for examaple, for the second post-treatment year (column (2)), by payment timing (columns (3) and (4)), and by type of tax (columns (5) and (6)). The estimates in all cases are close to zero, precisely estimated, and statistically insignificant.

One potential confounding factor for the lack of heterogeneity in signals on p and θ is that some subjects might have interpreted the *audit-statistics* message *per se* as a signal that their firms were on the IRS's radar, above and beyond the factual information conveyed in the message. We were careful to mitigate this concern in the design of our mailings by, for instance, underscoring in the letter that its recipients were randomly selected. Nevertheless, some individuals may have ignored or overlooked this cue. While recipients may have learned–or thought they learned–something from the receipt of the *audit-statistics* message, there is no reason why they could not also learn from the content of the message. In other words, the test presented above continues to be valid, and *A&S* would still predict that the *audit-statistics* message have a differential effect depending on the values of p and θ.

To address this concern more directly, we use the *audit-threat* treatment arm, where an explicit threat from the tax agency was made to every recipient. Panel (d) of Figure 4 depicts the difference in the evolution of VAT payments over time between the two sub-treatments in the *audit-threat* arm, corresponding to audit probabilities of 50% and 25%. We find almost no systematic difference between the two groups in post-treatment VAT payments. We can perform a more parametric test based on an econometric model similar to the one in equation (2) for firms assigned to the *audit-threat* letter:

$$Y_{it} = \alpha + \tau_p \cdot p_i + \delta \cdot Post_t + \gamma_p \left(p_i \cdot Post_t\right) + \epsilon_{it} \tag{3}$$

where $p_i \in \{0.25, 0.50\}$ is the audit probability in the *audit-threat* letter sent to firm i. Again, *A&S* predicts $\gamma_p > 0$. The results are presented in panel (b) of Table 3. While the estimated coefficient based on *audit-threat* messages implies an elasticity of 0.217 with respect to p in the first year post-treatment, this estimate is economically small and statistically insignificant (p-value=0.128). Taking into account precision and power concerns, the evidence from panel (d) of Figure 4 and panel (b) of Table 3 further reinforces our result. Contrary to *A&S* predictions, tax compliance does not seem to depend on the probability of being audited, even when there is a direct and credible threat of an audit, rather than simply information on audit probabilities.

The evidence is robust, then, that firms did not react to the values of p and θ shown in our letters.[58] Since a large portion of our subject pool was assigned this treatment arm, these elasticities are precisely estimated. It is not clear, however, whether the estimates are small enough, and precisely estimated enough, to rule out the values of the elasticities predicted by *A&S*. We address this question below.

5.2.2 *A&S* Calibration

For a quantitative test of *A&S*, we need quantitative predictions from *A&S*. In this section, we present results from different calibrations of the model and compare them to the experimental results in the following section.

Let Y be the total value-added amount and let $\tau = 0.22$ be the value added tax rate. Let E be the amount to be underreported (so $\tau \cdot E$ is the amount evaded). Each firm has a utility from income given by a Constant Relative Risk Aversion (CRRA) utility function with risk parameter σ. Let p be the probability that the tax return for a given year be audited sometime in the future, and θ the penalty rate applied over the amount evaded when caught (both of these parameters are defined as in the *audit-statistics* treatment).

Given any reasonable value for the CRRA parameter, the basic *A&S* model predicts 100% evasion. As a result, we need to use one of the extensions discussed in the literature to accommodate the 26% evasion rate observed in practice (as estimated by Gomez-Sabaini and Jimenez, 2012). We consider the following extensions: endogenous audit probabilities (Allingham and Sandmo, 1972; Yitzhaki, 1987), third-party reporting and whistle-blowing (Acemoglu and Jackson, 2017), misperceptions of audit parameters (Alm et al., 1992), and social preferences (Luttmer and Singhal, 2014).

The probability of being audited can be broken down as $p = p_0 + p_1 \frac{E}{Y}$.[59] The parameter $p_1 > 0$ represents the endogeneity of the audit process, whereby firms that evade more are more likely to be audited (in the original *A&S* model, the audit probability is exogenous so $p_1 = 0$). Firms can be caught evading by some non-audit technology such as third-party reporting or whistle-blowing. We represent this with an effective probability of being caught of $p + \epsilon$, where the parameter epsilon represents the additional monitoring tool. To allow for misperceptions, we can calibrate p and θ to the average perceptions reported in the survey

[58] See Appendix B.6.2 for a series of robustness checks (alternative specifications based on OLS, Tobit and Probit models, and using an alternative data source for the dependent variable), all of which yielded similar results. An additional robustness check, presented in the same Appendix, shows that results are robust if, instead of estimating elasticities with respect to p and θ separately, we estimate elasticity with respect to $p * \theta$ (i.e., the expected penalty per dollar evaded).

[59] We assume that, when an audit occurs, all evasion is detected. In practice, that probability may be smaller than one, which would only make the *A&S* result more puzzling: firms should be even less worried about being audited.

rather than the values calculated on the basis of the tax agency's administrative records. To allow for social preferences, we assume that individuals get some direct utility from paying taxes, and that that utility is equal to the fraction α of the amount paid. This social responsibility parameter α can take values from zero to one, where a higher value denotes a greater sense of social responsibility (in the original *A&S*, $\alpha = 0$).

The optimal evasion choice is given by maximizing the expected utility:

$$\max_{E\in[0,Y]} \frac{1-p\left(\frac{E}{Y}\right)-\epsilon}{1-\sigma}\Big(Y-\alpha\tau(Y-E)\Big)^{1-\sigma}+\frac{p\left(\frac{E}{Y}\right)+\epsilon}{1-\sigma}\Big(Y-\alpha\tau(Y-E)-(1+\theta)\tau E\Big)^{1-\sigma} \quad (4)$$

Given a set of parameters, it is straightforward to find the optimal value of E that solves this maximization problem.[60] Table 4 presents the calibration results; each row corresponds to a different calibration of *A&S*. The first seven columns correspond to the parameter values, while the last three indicate the corresponding predictions: $\frac{E}{Y}$ is the evasion rate, $\frac{\partial log(\tau(Y-E))}{\partial p}$ is the elasticity with respect to the audit probability, and $\frac{\partial log(\tau(Y-E))}{\partial \theta}$ is the elasticity with respect to the penalty rate.[61]

All the parameters are calibrated so that the predictions always match the average evasion rate ($\frac{E}{Y}$) of 26% (Gomez-Sabaini and Jimenez, 2012). In the first row, we assume a CRRA of 4 and set the audit probability and penalty rates at the rates estimated from administrative records ($p_0 = 0.117$ and $\theta = 0.306$). To match the 26% evasion rate, we allow for a non-audit detection rate of $\epsilon = 0.575$. The resulting elasticity is 4.516 with respect to p and 3.434 with respect to θ. This is the simplest extension to the *A&S* model, and given that its predictions are in the middle range of all our calibrations, it is our preferred specification. The remaining rows present results with alternative calibrations of the model. Even though the models are quite different, the predicted elasticities are in the same order of magnitude as our preferred specification.

In the second row, instead of accommodating the evasion rate of 26% by introducing the non-audit detection rate, we assume a social responsibility parameter of $\alpha = 0.202$. While this different approach yields different predicted elasticities, they are in the same order of magnitude: the elasticity with respect to the audit probability is 9.116, and the elasticity with respect to the penalty rate is 1.207. The specification in the third row is similar to the one in the second row, but it is augmented by allowing for an endogenous audit probability. We let $p_0 = p_1 = 0.0896$, which accommodates two important features of audit probabilities: the effective audit probability turns out to be equal to the observed average probability of

[60] For more details, see Appendix C.

[61] These two elasticities are defined exactly as in the econometric model from equations (2) and (3) to facilitate the comparison of the regression results and the calibrations.

11.7% and, consistent with the content of the *audit-endogeneity* message, a firm that does not evade taxes ($\frac{E}{Y} = 0$) would double its audit probability if it decided to evade taxes ($\frac{E}{Y} = 1$). Since this endogeneity parameter is not, on its own, enough to match the observed evasion rate, we again rely on the social responsibility parameter by setting $\alpha = 0.2296$. This specification shows that introducing endogeneity to the audit probabilities barely changes elasticity with respect to the audit probability (it is 3.324, similar to the 4.516 from the first specification), but it does substantially reduce elasticity with respect to the penalty rate (to 0.589). The fourth row follows a similar specification as the second row, but is extended by allowing individuals to have biased perceptions of the audits: $p_0 = 0.407$ and $\theta = 0.305$. These biases would not, on their own, be enough to match the observed evasion rate, so we set the social responsibility parameter to $\alpha = 0.643$. The elasticities yielded are, once again, of the same order of magnitude as with the other specifications: the elasticity is 3.889 with respect to audit probability and 1.763 with respect to the penalty rate. The specifications in the second set of four rows are identical to the ones in the first set of four rows, except the assumption of a CRRA of 2 instead of 4. The results indicate that assuming a lower level of risk aversion leads to elasticities that are even larger in magnitude.

5.2.3 Comparison of Experimental Results and the *A&S* Calibration

We can test the null hypothesis that the elasticities for p and θ in the main specification of the *audit-statistics* presented in column (1) of Table 3 ($\gamma_p = -0.063$ and $\gamma_\theta = -0.033$) are equal to the elasticities in our preferred *A&S* calibration. We can reject the null hypothesis that the elasticity is 4.516 for the audit probability and 3.434 for the penalty rate (both tests with p-values<0.001). In other words, the calibrated elasticities (4.516 for the audit probability and 3.434 for the penalty rate) far exceed the 95% confidence bands for the estimated elasticities ($[-0.536, 0.411]$ and γ_θ $[-0.264, 0.198]$, respectively).

One potential concern with the above comparison is the implicit assumption that a letter conveying the message of a one percentage point higher signal of p or θ will increase the recipient's perception of the parameter by one percentage point–and that is a lot to assume: some taxpayers may not have adjusted their prior belief all the way to the signal, and other taxpayers may not have even read the letter in the first place. As a benchmark, we can use the estimates from related studies that measure how individuals learn about economic variables such as the inflation rate (Cavallo et al., 2017), cost of living (Bottan and Perez-Truglia, 2017), and housing prices (Fuster et al., 2018). These studies find that for each percentage-point increase in the feedback given to subjects, the average individual alters their beliefs by about half a percentage point. If we assume this rate of learning, we should double the elasticities estimated in our regressions before comparing them to the calibrations of *A&S*.

Even under this assumption, we could still confidently reject the null hypothesis that the estimated elasticities are equal to the calibrated elasticities (p-values<0.001 for both γ_p and γ_θ).

The effect of the differential values of p and θ on the *audit-statistics* treatments, presented in Section 5.1 above, can provide a direct estimate of the learning rate in our context. The survey data suggest that a one percentage-point increase in the signal on audit probability in the letter increased the perceived audit probability nine months later by 0.397 percentage points (SE 0.288). Even if it is imprecisely estimated and thus statistically insignificant at conventional levels, this point estimate suggests a learning rate consistent with other studies of learning, which is reassuring. Moreover, because of multiple sources of noncompliance, that rate is probably less than the true learning rate.[62] We can, furthermore, reproduce the analysis with an extremely conservative assumption on the magnitude of the learning rate: even if we assumed that for each percentage-point difference in the letter individuals adjusted their beliefs by only one tenth of a percentage point, the null hypothesis that the estimated elasticities are equal to those in the *A&S* calibration (p-values of 0.033 and 0.002 for audit probability and for penalty rate, respectively) could be rejected.

5.2.4 Comparison to Related Studies

We can also compare our estimated elasticities to those from related studies. In some research, laboratory experiments are used to study tax evasion. These experiments often randomize the probability of being audited by experimenter and penalties involved. Consistent with our results, those laboratory studies find evidence of probability neglect. For example, Alm et al. (1992) find an elasticity of 0.169 with respect to audit probability (comparable to our estimate of -0.063), and an elasticity of 0.037 with respect to penalty rate (comparable to our estimate of -0.033). Indeed, these elasticities are statistically indistinguishable from the ones obtained in our study (p-values of the differences are 0.338 and 0.555 respectively).

We also compare our findings to the results of a couple of related field experiments. Dwenger et al. (2016) conducted a field experiment in the context of a local church tax in Germany for which enforcement was extremely lax. While this experiment was not designed to test the *A&S* model, it did include one treatment arm where the message mentioned different audit probabilities ($p = 0.1$, $p = 0.2$, or $p = 0.5$). The results are qualitatively consistent with our finding of probability neglect: the effects of all these probability messages are statistically indistinguishable from each other. Another related experiment, Kleven et al. (2011), included a treatment arm with two different audit probabilities. Consistent with our results, it finds rates of tax compliance between individuals assigned to different audit probabilities

[62] For a discussion of the sources of noncompliance, see footnote 55.

to be negligible in magnitude.[63] The evidence from Kleven et al. (2011), however, would be consistent with *A&S* because their subjects face automatic third-party reporting, which our subjects do not. The authors conducted their experiment with wage earners in a country where tax evasion is automatically detected through third-party reporting, regardless of audits. As a result, *A&S* predicts that, consistent with their evidence, wage earners would report their earnings truthfully regardless of probability of being audited.

6 Discussion: Risk-as-Feelings

In this section, we summarize our findings and discuss their potential interpretations and implications.

We present three main findings. First, we documented ***increased compliance***: on average, the *audit-statistics* message had a positive effect on tax compliance. Second, we reported ***reduced subjective probability***: on average, the *audit-statistics* message decreased the perceived probability of being audited. Third, we documented ***probability neglect***: the effect of the *audit-statistics* message did not depend on the audit probability conveyed in the letter or on the firm's prior belief about that probability. Jointly, these three findings are inconsistent with the *A&S* predictions. But the question of what framework might provide a better fit for these results remains unanswered.

One natural candidate is salience (Chetty et al., 2009): firms may behave as if the probability of detection and the penalty rate were zero unless those parameters are made salient to them. This explanation could reconcile the findings of increased compliance and reduced perceived audit probability: even if firms who were sent the messages adjusted downward their perceived probability of audit, they would have behaved as if that probability were zero had they not received those signals. However, the salience model fails to explain other features of our findings. First, salience effects are by definition short-lived. A reminder of a non-salient tax should only affect the behavior of an agent at the time the information is received, not days or months later. Effectively, salience models predict a rapid decay of the effect of information, which contradicts our evidence that the *audit-statistics* letter had effects that persisted months after the information was transmitted.[64] The salience model is

[63]In one of their treatments, they send letters to individuals stating audit probabilities as high as $p = 50\%$ and $p = 100\%$. Compared with a group that did not receive any letter, they find that the letters had a positive and significant effect on declared income and tax liability. While statistically significant, the differential effect between these two groups is economically negligible: an increase from 50% to 100% in the signal on the probability of audit increases reported income by 0.025% and taxes paid by 0.05%.

[64]The informational treatment may increase salience and cause an instantaneous effect with lasting consequences if, for instance, it induces a change in the way the firm deals with evasion in transactions. Such a change would imply a constant effect over time, however, and we find a substantial decline in evasion over

also inconsistent with our finding of *probability neglect:* according to that model, the effect of making salient a high audit probability should be greater than the effect of making salient a low audit probability.[65]

A second explanation would be agency issues at the firm if, say, the person who received the letter was not the person who decided how much tax to evade. This type of agency issue would generate insensitivity to the information received (or, at least, an attenuation effect). We find, however, that firms do react to the information received, just not in the direction predicted by *A&S* (i.e., information on low audit probabilities reduces, rather than increases, evasion). Moreover, agency and information frictions should be weaker in smaller firms, and the firms in our experimental sample are small: over 75% of firms in our sample have five or fewer employees.[66] Moreover, the heterogeneity analysis indicates no substantial or statistically significant difference between firms below and above the median number of employees (one to three employees versus the rest), which further reinforces the point that agency issues are not a decisive factor in our context.[67]

Our preferred interpretation is based on the model of risk-as-feelings (Loewenstein et al., 2001). The traditional economic models used for choice under uncertainty are cognitive in that agents make decisions using some sort of expectation-based calculus. The risk-as-feelings model proposes that responses to fearsome situations may differ substantially from cognitive evaluations of the same situations (Loewenstein and Lerner, 2003).[68] When fear is involved, agents tend to neglect the cost-benefit calculus and instead have quick, automatic, and intuitive responses to risk. A key prediction of this model is that feelings about risk are largely insensitive to changes in probability, in what the literature calls probability neglect (Sunstein, 2002; Zeckhauser and Sunstein, 2010). According to this model, the emotional response to risk makes individuals ignore the underlying likelihoods. There is evidence of probability neglect in a range of fearsome situations involving electric shocks, arsenic, abandoned hazardous waste dumps, pesticides, and anthrax (Sunstein, 2003; Zeckhauser and Sunstein, 2010).

This risk-as-feelings model can reconcile our three key findings, namely *increased compli-*

---

the year following the intervention.

[65]Another pertinent model from behavioral economics is prospect theory (Kahneman and Tversky, 1979; see, for example, Dhami and al Nowaihi, 2007). This model, however, is unlikely to explain our findings on, for instance, *probability neglect*: although differences between extremely low probabilities can be ignored under prospect theory, the range of probabilities in our context (e.g., in the *audit-threat* arm, the probabilities were 25% versus 50%) was far higher than what is normally considered extremely low in prospect theory.

[66]More precisely, 29.1% of the firms have a single employee, 46.2% have between two and five employees, and 15% have between six and ten employees.

[67]See Appendix B.8.

[68]A related concept, the affect heuristic, corresponds to quick, automatic, and intuitive evaluations of risky situations based on emotions, which might be used as a shortcut for more complex evaluations of risk (Slovic et al., 2004). Borrowing Kahneman's (2003) terminology for the dual system model of the human mind, emotions might influence the intuitive system.

ance and *reduced subjective probability*: even if the perceived probability of audit decreased among the treated subjects, they may nonetheless be scared into paying more taxes because they did not rely on cognitive evaluations of probabilities. The risk-as-feelings model predicts *probability neglect* and thus fits our third finding too.

The risk-as-feeling model suggests that taxpayers overreact to the threat of audits.[69] This interpretation can explain the paradox that, despite low audit probabilities and penalty rates, most taxpayers report the threat of audits as a major reason for reporting taxable income truthfully. For example, a survey by the United States Internal Revenue Service (2018) indicates that 61% of U.S. taxpayers claim that a "fear of audits" exerts a significant influence on their tax compliance decisions.[70] In comparison, audits are perceived as a strong a deterrent as third-party reporting: 66% of respondents identified "third-party reporting (e.g., wages, interest, dividends)" as an important factor for tax compliance. Moreover, there is some direct evidence that, consistent with the risk-as-feelings model, taxpayers have an emotional reaction to the thought of tax audits and the tax authority more generally. Some of the evidence comes from laboratory experiments. Coricelli et al. (2010), for instance, conducted a tax evasion game in the laboratory and measured how emotional arousal affected tax evasion decisions. They showed that the intensity of emotional arousal predicts whether and how much individuals evade. In a related laboratory study, Dulleck et al. (2016) showed a significant correlation between tax compliance and physiological markers of stress in making decisions about tax filing. Fear of tax audits can also be found in the media. For example, a Washington Post (2016) article claims that "a lot of people are super scared of the Internal Revenue Service" and that its powers "can instill a lot of fear." The New York Times (2009) reported cases where fear of the tax authority is strong enough to be considered a phobia.

In other areas of public policy, the risk-as-feelings heuristic can be a problem. It distorts facts and leads to irrational judgment, which results in suboptimal decisions from a pure risk-assessment perspective. Zeckhauser and Sunstein (2010) and others discuss cases involving regulation of nuclear power, vaccines, and other heated emotional issues. When it comes to tax collection, such behavioral biases might have positive implications for the tax authority's goals. Indeed, there is anecdotal evidence that tax authorities use fear tactics to foster tax compliance. In the United States, for example, a disproportionately large number of tax enforcement press releases covering criminal convictions and civil injunctions are re-

[69]This excessive caution has been documented in other contexts (Loewenstein et al., 2001). A fear of terrorist attacks, for instance, can make people choose not to take airplanes but, rather, other, more dangerous, forms of transport; and a fear of shark attacks can lead to unnecessary legislation (Sunstein, 2002, 2003; Zeckhauser and Sunstein, 2010).

[70]More precisely, 32% of respondents claim that a "fear of audits" exerts "a great deal of an influence," and 29% "somewhat of an influence" on whether they honestly report and pay their taxes.

leased during the weeks immediately preceding Tax Day, presumably to scare taxpayers into preparing compliant returns (Morse, 2009; Blank and Levin, 2010). Some tax experts even claim that the IRS "likes [targeting] celebrities because they get the most bang for their buck in terms of publicity" to "scare the public into complying" (Forbes, 2008).

The risk-as-feelings framework indicates that vivid imagery can be used to instill fear and biased risk evaluations (Slovic et al., 2004; Zeckhauser and Sunstein, 2010). Coincidentally, tax agencies seem to resort to vivid images in some of their advertising campaigns. A TV advertisement in the United States showed the IRS as "something like poltergeist coming out of a TV set and the world falling apart," followed by the phrase, "Have you filed your income tax?" (United Press International, 1988) The U.K. tax agency used advertisement campaigns that also rely on frightening imagery. One poster features a pair of eyes peeking threateningly through a gash in a piece of paper that reads, "If you've declared all your income, you have nothing to fear."[71] This anecdotal evidence suggests that some tax administrations may be leveraging fear to help in collecting taxes.[72]

7 Conclusions

The canonical model Allingham and Sandmo (1972) predicts that firms evade taxes by making optimal trade-offs on the costs and benefits of evasion. It is unclear, however, whether real-world firms react to audits according to that model. We designed a large-scale field experiment in collaboration with Uruguay's tax authority to assess the factors behind firms' evasion behavior and reactions to audits. Our findings indicate that firms do increase tax compliance when informed of the audit process. We do not, however, find this reaction to be consistent with the predictions of *A&S*. For example, the information on audits decreased (rather than increased) the perceived probability of being audited; moreover, the effects of our messages about audit probabilities were independent from the signal we conveyed and from the firms' prior beliefs. Models of salience are consistent with the increased compliance and reduced perceived perception of audit probabilities that we observed, but they are not consistent with our findings on probability neglect. All three findings can, we agree, be reconciled by the risk-as-feelings model, which heeds the role of emotions in decision-making and predicts that agents tend to exhibit probability neglect in dreaded or feared situations,

[71] This poster is reproduced in Appendix D.

[72] Whether these fear tactics should be used by tax agencies is outside the scope of this paper. These tactics may be ethically questionable to the extent to which they rely on deception. Moreover, actively promoting fear could have unintended negative effects, such as imposing negative psychological stress on taxpayers. For a discussion of the ethical and practical issues at stake in communication efforts to increase tax compliance, see, for example, Morse (2009).

like paying taxes.

Our findings also contribute to the more general debate on the determinants of tax compliance. One of the main puzzles in the literature is overly low evasion rates . Third-party reporting can explain high levels of compliance for some sources of income, such as wage income (Kleven et al., 2011). We would expect, however, much higher evasion rates in other contexts, such as for self-employed income, where there is limited third-party reporting and low detection probabilities and penalty rates. One traditional explanation for this phenomenon is tax morale: firms and individuals do not evade taxes because complying with tax obligations is the right thing to do (Luttmer and Singhal, 2014). Our evidence suggests an alternative explanation: because tax decisions are emotional, audits scare taxpayers into compliance just as scarecrows scare birds.

We conclude by discussing some policy implications. In the traditional *A&S* framework, the relevant policy lever is the number of audits: the tax agency must find the point at which the marginal cost of an additional audit equals the expected marginal benefit (i.e., higher tax revenues). Our findings suggest that small and medium-sized firms face significant information and optimization frictions when reacting to audits. These frictions introduce new levers for policy-making. Tax agencies can, for instance, decide whether to be transparent about the audit process,[73] whether to contact taxpayers to remind them of it, and whether to make the costs of being a tax cheat salient and vivid through advertisement campaigns.[74] There is anecdotal evidence that some tax agencies already have working knowledge of these policy levers. Some tax agencies seem to avoid transparency about the auditing process while increasing the visibility of enforcement actions around tax day. Some even make direct reference to fear in their advertisement campaigns. There is no direct evidence, though, on whether these policies effectively increase tax compliance or whether they have unintended effects, such as instigating so much fear in taxpayers that their anxiety and unhappiness trump the positive effects of increased tax revenues. As stated by Alm (2019) in a recent review of the literature, "the role of emotions in tax compliance decisions remains largely unexamined." Our results highlight the need for more research on probability neglect in the decision to comply with tax obligations. Additional research should examine the role emotions play on other important economic choices, not only tax compliance.

[73]On the one hand, our evidence indicates that increasing transparency about audit probability will reduce the average perceived probability of being audited, which could reduce tax compliance. On the other, our finding of probability neglect suggests that, in the end, the reduction in perceived audit probability may not affect tax compliance.

[74]For a practical discussion of how to implement this type of policy, and the drawbacks, see Morse (2009). This same principle can be used to increase compliance with other laws. Dur and Vollaard (2019) provide experimental evidence to show that the salience of law enforcement can be used to reduce illegal garbage disposal, for instance.

References

Acemoglu, D. and M. O. Jackson (2017). Social Norms and the Enforcement of Laws. *Journal of the European Economic Association 15*(2), 245–295.

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

Alm, J. (2019). What motivates tax compliance? *Journal of Economic Surveys 33*(2), 353–388.

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.

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? *Virginia Tax Review 30*.

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.

Bott, K. M., A. W. Cappelen, E. Ø. Sørensen, and B. Tungodden (2020). You've got mail: A randomized field experiment on tax evasion. *Management Science 66*(7), 2801–2819.

Bottan, N. L. and R. Perez-Truglia (2017). Choosing Your Pond: Location Choices and Relative Income. *NBER Working Paper* (23615).

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.

Bruine de Bruin, W., 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.

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.

Cavallo, A., G. Cruces, and R. Perez-Truglia (2017, July). Inflation Expectations, Learning, and Supermarket Prices: Evidence from Survey Experiments. *American Economic Journal: Macroeconomics 9*(3), 1–35.

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

Coricelli, G., M. Joffily, C. Montmarquette, and M. C. Villeval (2010). Cheating, Emotions, and Rationality: An Experiment on Tax Evasion. *Experimental Economics 13*(2), 226–247.

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.

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.

Dulleck, U., J. Fooken, C. Newton, A. Ristl, M. Schaffner, and B. Torgler (2016). Tax compliance and psychic costs: Behavioral experimental evidence using a physiological marker. *Journal of Public Economics 134*, 9 – 18.

Dur, R. and B. Vollaard (2019). Salience of law enforcement a field experiment. *Journal of Environmental Economics and Management 93*(C), 208–220.

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.

Forbes (2008, July). Pity The Celebrity Taxpayer. *Forbes*.

Fuster, A., R. Perez-Truglia, and B. Zafar (2018). Expectations with Endogenous Information Acquisition: An Experimental Investigation. *NBER Working Paper*.

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. (2003). A Perspective on Judgement and Choice: Mapping Bounded Rationality. *American Psychologist 58*(9), 697–720.

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.

Kleven, H. J., C. Kreiner, and E. Saez (2016). Why Can Modern Governments Tax So Much? An Agency Model of Firms as Fiscal Intermediaries. *Economica 83*, 2016.

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. and S. Lerner (2003). The role of affect in decision making. In R. Davidson, K. Scherer, and H. Goldsmith (Eds.), *Handbook of Affictive Sciences*. Oxford: Oxford University Press.

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.

Meiselman, B. S. (2018). Ghostbusting in detroit: Evidence on nonfilers from a controlled field experiment. *Journal of Public Economics 158*, 180 – 193.

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

Naritomi, J. (2019, September). Consumers as tax auditors.

Nathan, B., R. Perez-Truglia, and A. Zentner (2020). My taxes are too darn high: Tax protests as revealed preferences for redistribution. *NBER Working Paper No. 27816*.

New York Times (2009, April). A Paralyzing Fear of Filing Taxes. *New York Times*.

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.

Reinganum, J. F. and L. L. Wilde (1985). Income tax compliance in a principal agent framework. *Journal of Public Economics 26*(1), 1–18.

Saez, E., J. Slemrod, and S. Giertz (2012). The elasticity of taxable income with respect to marginal tax rates: A critical review. *Journal of Economic Literature 50*(1), 3–50.

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.

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.

Srinivasan, T. N. (1973). Tax Evasion: A Model. *Journal of Public Economics 2*(4), 339–346.

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

Sunstein, C. R. (2003). Terrorism and probability neglect. *Journal of Risk and Uncertainty 26*(2-3), 121–136.

United Press International (1988). Psychologist takes issue with irs scare tactic. *UPI-United Press International*.

United States Internal Revenue Service (2018). Comprehensive Taxpayer Attitude Survey (CTAS) 2017 Executive Report. Publication 5296 (Rev. 3-2018) Catalog Number 71353Y, Department of Treasury, Washington, D.C.

Washington Post (2016, August). That is NOT the IRS Calling You! *The Washington Post*.

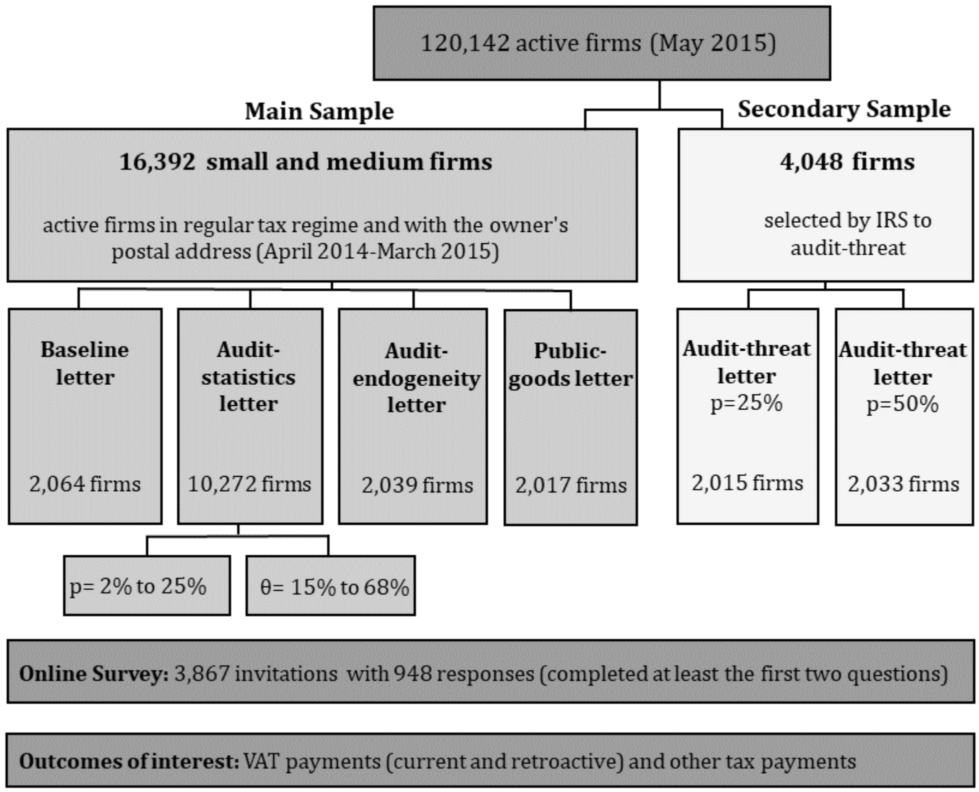
Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data.* MIT press.

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

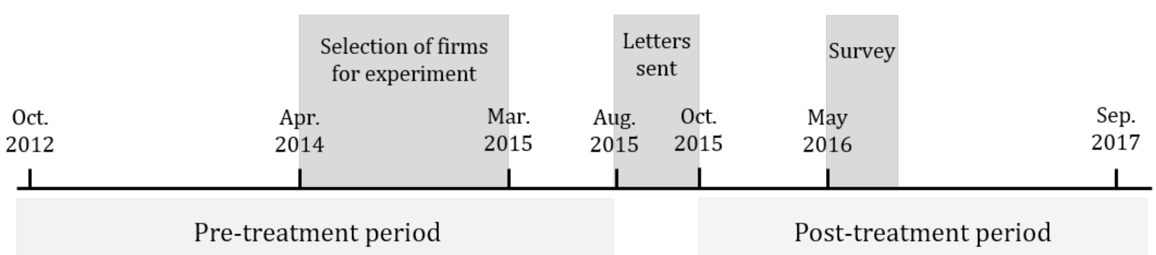
Zeckhauser, R. and C. R. Sunstein (2010). Dreadful Possibilities, Neglected Probabilities. In E. Michel-Kerjan and P. Slovic (Eds.), *The Irrational Economist: Making Decisions in a Dangerous World*, pp. 116–123. New York: Public Affairs Press.

Figure 1: Structure of the Field Experiment

a. Samples and Treatment Arms



b. Timeline



Notes: Panel (a) reports the key features of the experimental design. Panel (b) reports the key dates of the field experiment and the survey.

Figure 2: Sample 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.

MESSAGE

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,

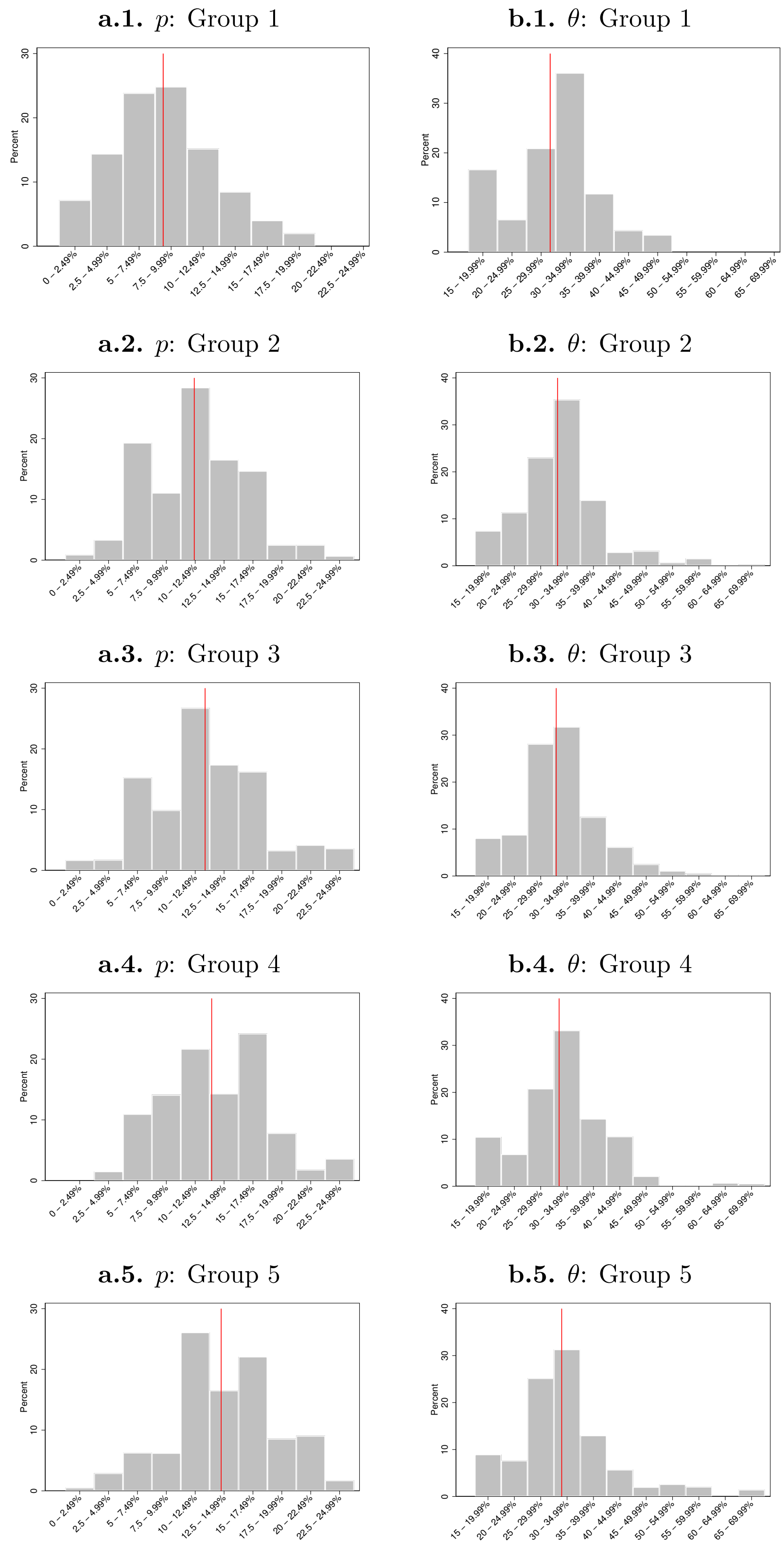
El Director General de Rentas

Lic. Joaquín Serra

Collection and Controls Division
Internal Revenues Services

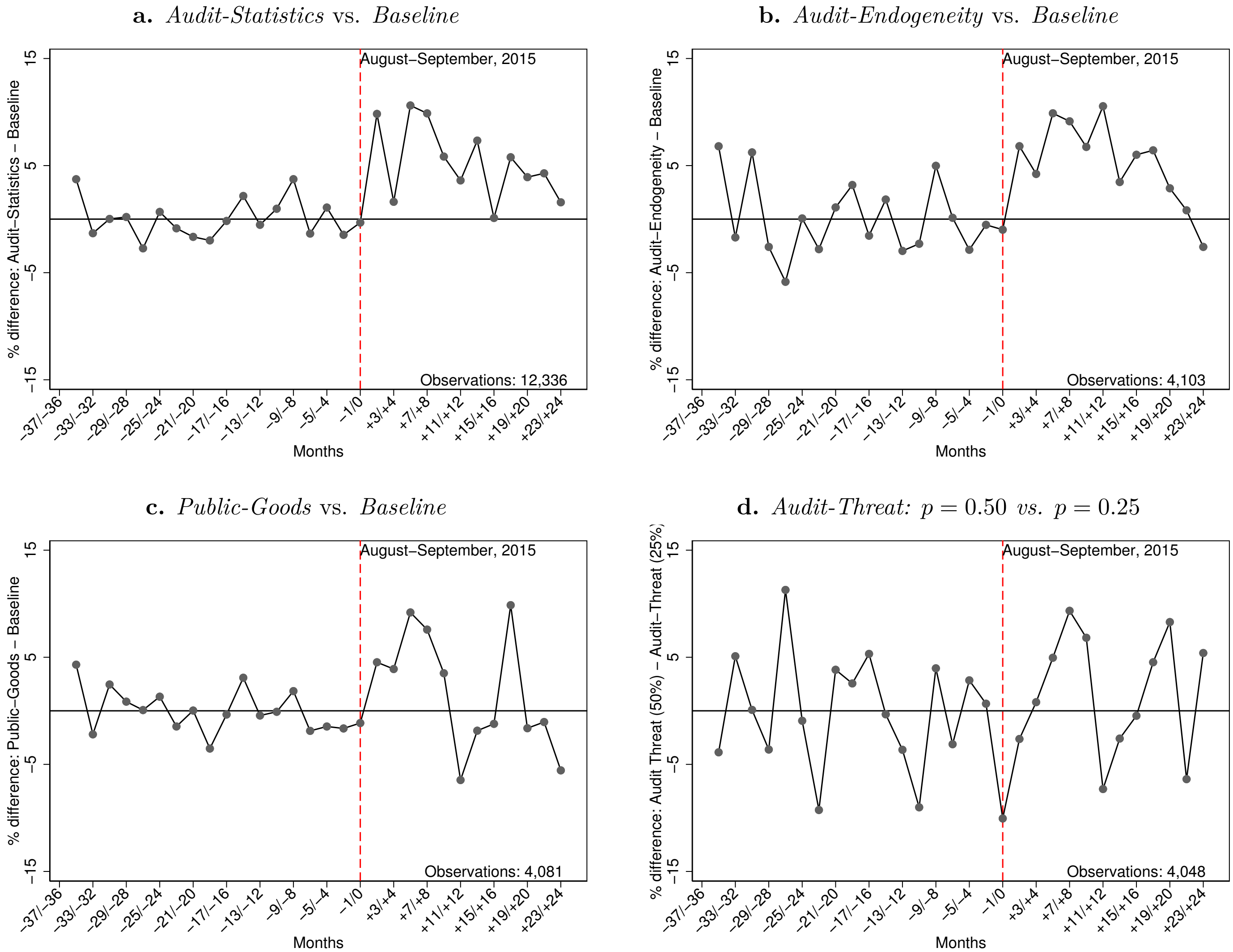
Notes: The *baseline* letter contains information on the goals and responsibilities of the tax authority. In the space with the text MESSAGE, the baseline letter is empty (See A.1 for the full letter). In the *audit-statistics* letter, a paragraph added to the baseline letter provides information on audit probabilities and tax evasion penalty rates (Appendix A.2). In the *audit-threat* letter, firms were randomly assigned to groups with different probabilities (25% and 50%) of being audited in the following year (Appendix A.3). The *audit-endogeneity* letter included information on how evading taxes typically doubles the probability of being audited (Appendix A.4). Finally, the *public-goods* letter included a message with information on the cost of evasion in terms of the provision of public goods (Appendix A.5).

Figure 3: Distribution of Statistics Shown in *Audit-Statistics* Letters by VAT Payment Quintiles



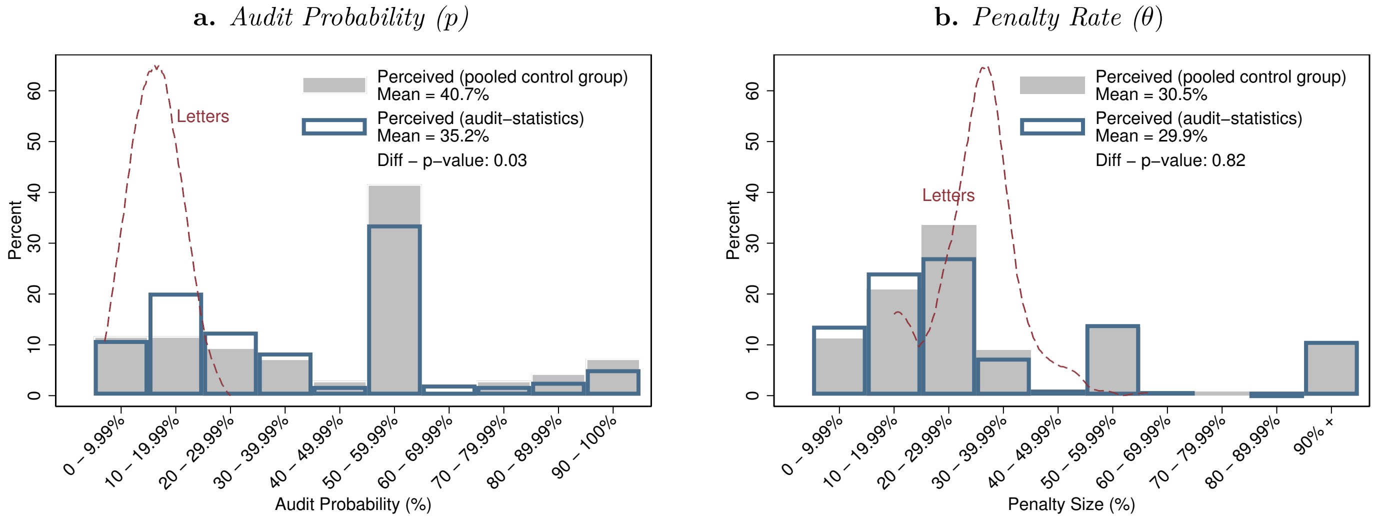
Notes: N=10,272. These panels show the information provided in the *audit-statistics* letter, including the probability of being audited (p in panel (a)) and the penalty rate (θ in panel (b)). Groups one through five correspond to each of the pre-treatment VAT payment quintiles (group one being the bottom quintile and group five being the top quintile). In each panel, the red vertical line denotes the average audit probability or penalty rate for all the members of the group.

Figure 4: Effects of *Audit-Statistics*, *Audit-Endogeneity*, *Public-Goods*, and *Audit-Threat* Messages on VAT Payments



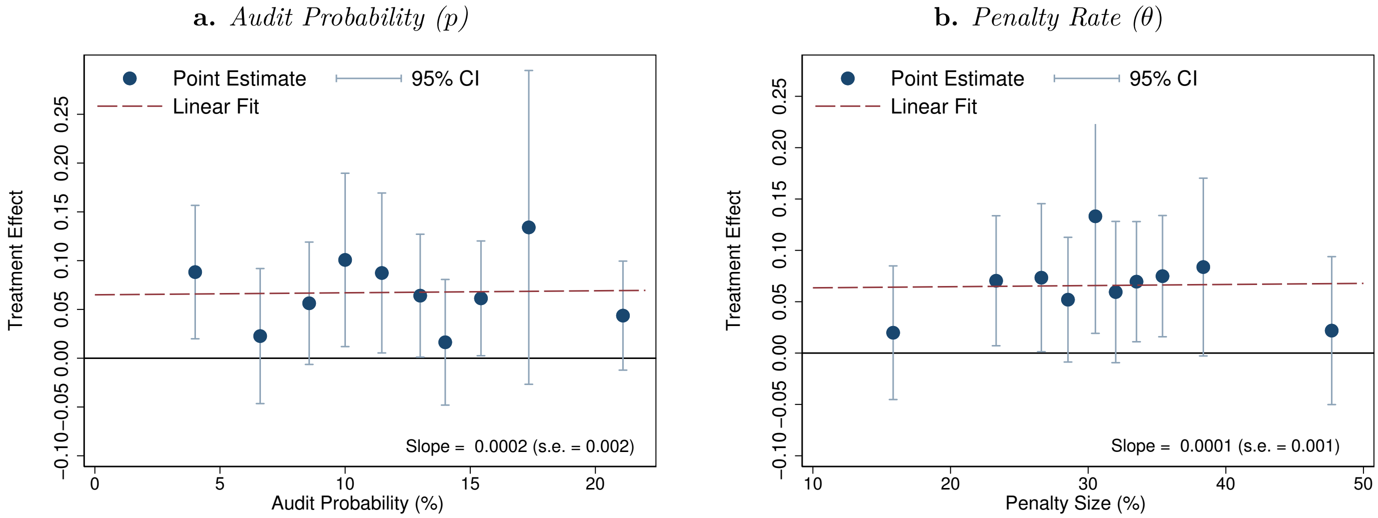
Notes: These figures plot the percentage difference in bimonthly total VAT payments between treatment and control groups, normalized by the average pre-treatment percentage difference (i.e. between months -35 and 0) for the same outcome. The data cover the period from October 2012 to September 2017. The months of August and September 2015–when most of the letters were delivered–are defined as the reference bimonthly period (and marked with the dashed vertical line). Panel (a) presents the effect of the *audit-statistics* message (i.e., the difference between *audit-statistics* and *baseline* letters), while panel (b) represents the effect of the *audit-endogeneity* message and panel (c) depicts the effect of the *public-goods* message. Panel (d) presents the difference between being assigned a 50% probability of being audited (p =50%) and a 25% probability of being audited (p =25%) in the *audit-threat* letters. For each pair of months, VAT payments are top-coded at the 99.99% percentile to avoid contamination of the results by outliers.

Figure 5: Survey Results: Perception of Audit Probabilities and of Tax Evasion Penalty Rates by Treatment Group



Notes: The histograms are based on the survey responses of individuals who did not self-identify as non-owners. *Perceived (pooled control group)* (N=137) refers to survey respondents who received the *baseline* (N=69) or the *public-goods* (N=68) letters during the experimental stage (neither of those letters contained any information on audit probabilities or penalty rates). *Perceived (audit–statistics)* refers to respondents who received the *audit-statistics* letters (N=365). In panel (a), the x-axis represents the probability of being audited; in panel (b), it represents the average penalty rate. We report the mean responses and the p-value of the difference between the two groups. The answers correspond to Q2 and Q4 in the survey (see full survey questionnaire in Appendix A.7). The red line represents the density function of the information displayed in the *audit-statistics* letters, measured in the right y-axis (hidden for the sake clarity).

Figure 6: Effect of *Audit-Statistics* vs. *Baseline* by Deciles of p and θ



Notes: Panel (a) plots the effect of the *audit-statistics* letter on total VAT payments by decile of p in the first year post-treatment (October 2015–September 2016), while panel (b) reports the results from the same regressions by decile of θ (N=10,272). In both panels, each dot represents the estimated treatment effect for each decile of the parameter considered. These effects are estimated using a regression similar to the one reported in equation (1), but with two differences. First, instead of including a single treatment variable, we include ten dummy variables, one for each decile of p or θ. These dummies take the value of one if the signal in the letter belongs to the corresponding decile in the p or θ distribution, and zero if the signal corresponds to a different decile, or if the firm was assigned to the *baseline* treatment. Second, we include an additional set of dummies for quintiles of the pre-treatment VAT payments, which are the groups from which we drew the sample of "similar firms" to calculate p and θ and the corresponding interactions with the post-treatment indicator. All effects are depicted with a 95% confidence interval. The results are based on Poisson regressions, so the coefficients can be interpreted directly as semi-elasticities. Confidence intervals are computed with standard errors clustered at the firm level. The dashed line represents the linear fit that results from regressing the treatment effect on the average signal within the decile.

Table 1: Balance of 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 (3 months pre-mailing)	0.925	0.939	0.926	0.928	0.181	0.897	0.891	0.538
	(0.003)	(0.005)	(0.006)	(0.006)		(0.007)	(0.007)	
Amount of VAT paid (3 months pre-mailing)	1.872	1.963	1.926	1.906	0.557	1.739	1.748	0.950
	(0.027)	(0.067)	(0.069)	(0.059)		(0.097)	(0.092)	
Years registered with tax agency	15.338	14.746	15.704	15.009	0.268	19.453	19.425	0.944
	(0.170)	(0.224)	(0.538)	(0.225)		(0.285)	(0.286)	
Share audited between 2013-2015	0.106	0.097	0.089	0.101	0.302	0.134	0.147	0.382
	(0.004)	(0.009)	(0.009)	(0.009)		(0.010)	(0.010)	
Number of employees	4.814	4.658	4.880	5.089	0.962	4.835	4.880	0.795
	(0.264)	(0.538)	(0.566)	(0.635)		(0.126)	(0.117)	
Share filed comprehensive tax return in 2013	0.682	0.687	0.691	0.687	0.871	0.999	1.000	0.558
	(0.005)	(0.010)	(0.010)	(0.010)		(0.001)	(0.000)	
Share no retail goods sector	0.289	0.293	0.283	0.300	0.621	0.431	0.434	0.845
	(0.004)	(0.010)	(0.010)	(0.010)		(0.011)	(0.011)	
Share retail goods sector	0.218	0.219	0.214	0.227	0.775	0.334	0.322	0.398
	(0.004)	(0.009)	(0.009)	(0.009)		(0.011)	(0.010)	
Share services sector	0.493	0.488	0.504	0.473	0.232	0.235	0.244	0.482
	(0.005)	(0.011)	(0.011)	(0.011)		(0.009)	(0.010)	
N	10,272	2,017	2,039	2,064		2,015	2,033	

Notes: Averages for different pre-treatment firm-level characteristics, disaggregated by treatment group and type of sample (robust standard errors are reported in parentheses). The main sample includes all firms selected as described in section 3.2. The secondary sample includes high-risk firms selected by the IRS for the *audit-threat* treatment. 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 the VAT amount and firm characteristics come from administrative tax records (including monthly payments, annual tax returns, and auditing registers). The amount of VAT reported in row 2 is expressed in constant thousands of U.S. dollars as of August 2015.

Table 2: Average Effects of *Audit-Statistics*, *Audit-Endogeneity,* and *Public-Goods* Messages on VAT and Other Tax Payments by Time Horizon and Payment Timing

	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* (10,272 firms) vs *Baseline* (2,064 firms)						
Post-Treatment	0.070***	0.032	0.383***	0.053**	0.086**	0.073***
	(0.021)	(0.027)	(0.140)	(0.021)	(0.037)	(0.020)
Pre-Treatment	0.009	0.004	-0.048	0.012	0.008	0.014
	(0.020)	(0.026)	(0.118)	(0.020)	(0.043)	(0.021)
b. *Audit-Endogeneity* (2,039 firms) vs *Baseline* (2,064 firms)						
Post-Treatment	0.071***	0.032	0.264*	0.061**	0.090*	0.078***
	(0.028)	(0.036)	(0.160)	(0.028)	(0.054)	(0.028)
Pre-Treatment	-0.005	-0.009	0.097	-0.010	0.056	0.017
	(0.028)	(0.035)	(0.164)	(0.028)	(0.055)	(0.028)
c. *Public-Goods* (2,017 firms) vs *Baseline* (2,064 firms)						
Post-Treatment	0.051**	0.004	0.208	0.043*	0.067	0.056**
	(0.025)	(0.032)	(0.170)	(0.025)	(0.043)	(0.024)
Pre-Treatment	-0.003	-0.017	-0.088	0.001	-0.038	-0.015
	(0.024)	(0.033)	(0.163)	(0.024)	(0.054)	(0.026)

Notes: * significant at the 10% level, ** at the 5% level, *** at the 1% level. Standard errors reported in parentheses are clustered at the firm level. Treatment effects are estimated using the difference-in-differences specification reported in equation (1), which compares treated firms to control firms and pre-treatment to post-treatment periods using yearly aggregated variables. The results are based on Poisson regressions, so the 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 comparison for the *audit-endogeneity* and *public-goods* messages respectively. In the first row of each panel ("Post-Treatment"), the coefficient reported corresponds to a comparison between a post-treatment period and a pre-treatment period. The second row ("Pre-Treatment") presents a falsification test where two pre-treatment periods are compared. Columns (1) and (2) report the effect of treatment by time horizon. The post-treatment effect reported in column (1) corresponds to the difference-in-differences estimate that compares October 2015–September 2016 to October 2014–September 2015. The post-treatment effect reported in column (2) is analogous but uses the second year after the treatment as the post-treatment period (i.e., October 2016–September 2017). For the falsification tests, column (1) is based on a comparison between October 2014–September 2015 and October 2013–September 2014, while column (2) compares October 2014–September 2015 to October 2012–September 2013. Columns (3) and (4) present the first-year effect of treatment on retroactive (3) and concurrent (4) VAT payments. Columns (5) and (6) report the first-year results by type of tax. Column (5) presents the effect of the treatment on other (non-VAT) tax payments, while column (6) reports the effect on the total amount of taxes paid by the firms during the same period. In all cases, we restrict the analysis to firms that effectively received the letter as reported by the postal service.

Table 3: Elasticities of Tax Payments 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* (10,272 firms)						
Audit Probability (%)						
Post-Treatment	-0.063	0.076	0.009	-0.040	0.109	0.038
	(0.242)	(0.232)	(1.103)	(0.249)	(0.240)	(0.208)
Pre-Treatment	0.141	0.018	-1.709	0.229	-0.035	0.063
	(0.164)	(0.203)	(1.118)	(0.162)	(0.230)	(0.147)
Penalty Size (%)						
Post-Treatment	-0.033	-0.175	0.928	-0.098	0.061	-0.001
	(0.118)	(0.134)	(0.763)	(0.114)	(0.103)	(0.092)
Pre-Treatment	-0.128	-0.163	0.204	-0.145	0.018	-0.078
	(0.108)	(0.127)	(0.524)	(0.111)	(0.119)	(0.087)
b. *Audit-Threat* (4,048 firms)						
Audit Probability (%)						
Post-Treatment	0.217	0.250	-0.347	0.205	0.002	0.233**
	(0.142)	(0.175)	(0.676)	(0.209)	(0.176)	(0.111)
Pre-Treatment	-0.185	-0.193	-0.432	-0.149	-0.067	-0.257
	(0.157)	(0.171)	(0.676)	(0.125)	(0.148)	(0.164)

Notes: * significant at the 10% level, ** at the 5% level, *** at the 1% level. Standard errors reported in parentheses are clustered at the firm level. Treatment effects are estimated using the difference-in-differences specification reported in equation (2) which compares treated firms that received different signals on p and θ. In all cases, we include an additional set of dummies for quintiles of the pre-treatment VAT payments, which are the groups from which we drew the sample of "similar firms" to calculate p and θ and the corresponding interactions with the time variable. The results are based on Poisson regressions with variables expressed in percentage terms, so the coefficients can be interpreted directly as elasticities. 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. For example, rows (1) and (3) of panel (a) present the effect of an additional percentage point of p and θ (respectively) in the information included in the letters on post-treatment VAT payments. In the "Post-Treatment" rows, the coefficient reported corresponds to a comparison of a post-treatment period and a pre-treatment period. In the "Pre-Treatment" rows we present a falsification test where two pre-treatment periods are compared. Columns (1) and (2) report the effect of treatment by time horizon. The post-treatment effect reported in column (1) corresponds to the difference-in-differences estimate that compares October 2015–September 2016 to October 2014–September 2015. The post-treatment effect reported in column (2) is analogous but uses the second year after the treatment as the post-treatment period (i.e., October 2016–September 2017). For the falsification tests, column (1) is based on a comparison between October 2014–September 2015 and October 2013–September 2014, while column (2) compares October 2014–September 2015 to October 2012–September 2013. Columns (3) and (4) present the firstyear effect of treatment on retroactive (3) and concurrent (4) VAT payments. Columns (5) and (6) report the first-year results by type of tax. Column (5) presents the effect of the treatment on other (non-VAT) tax payments, while column (6) reports the effect on the total amount of taxes paid by the firms during the same period. In all cases, we restrict the analysis to firms that effectively received the letter as reported by the postal service.

Table 4: Predicted Elasticities under Different *A&S* Calibrations

Setup							Predictions		
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	0.117	0	0.575	0.306	1	0.26	4.516	3.434
4	0.22	0.117	0	0	0.306	0.202	0.26	9.116	1.207
4	0.22	0.0896	0.0896	0	0.306	0.2296	0.26	3.324	0.589
4	0.22	0.407	0	0	0.305	0.643	0.26	3.889	1.763
2	0.22	0.117	0	0.614	0.306	1	0.26	9.777	6.578
2	0.22	0.117	0	0	0.306	0.176	0.26	18.245	2.111
2	0.22	0.0896	0.0896	0	0.306	0.2022	0.26	4.215	0.661
2	0.22	0.407	0	0	0.305	0.586	0.26	7.771	3.030

Notes: Each row corresponds to a different calibration of the extended *A&S* model presented in Section 5.2.2. 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 that rate.

Online Appendix: For Online Publication Only

Tax Audits as Scarecrows: Evidence from a Large-Scale Field Experiment

Marcelo Bergolo, Rodrigo Ceni, Guillermo Cruces, Matias Giaccobasso, and Ricardo Perez-Truglia

04/11/21

A Letters and Survey

This appendix presents samples of the five types of letters sent in our experiment: the *baseline* letter (A.1), the *audit-statistics* letter (A.2), the *audit-threat* letter (A.3), the *audit-endogeneity* letter (A.4), and the *public goods* letter (A.5). Appendix A.6 presents a sample of the email invitation to complete the online survey sent by the IRS, and Appendix A.7 presents the questionnaire module about perceptions of audit probabilities and penalty rates included in the survey.

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.

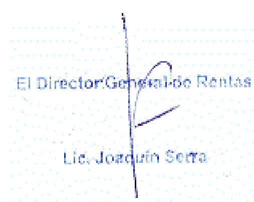
On the basis of historical information on similar businesses, there is a probability of p% that the tax returns you filed for this year will be audited in at least one of the coming three years. If, pursuant to that auditing, it is determined that tax evasion has occurred, you will be required to pay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



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 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 (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 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.

□Somewhat sure.

□ Very sure.

Q6) 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?

□ It would increase significantly.

□It would increase slightly.

□It would not change.

□It would diminish slightly.

□It would diminish significantly.

B Additional Results, Specifications, and Robustness Checks

B.1 Descriptive Statistics

Table B.1 reports firms' characteristics including VAT payments made in the three months before we sent the letters, the number of years each firm was registered with the IRS, the number of employees, and other basic variables. Column (2) provides these statistics for all firms in the main experimental sample. On average, firms in our sample had 4.8 employees and had been registered with the IRS for 15.2 years. Ten percent of the firms had been audited at least once sometime in the previous three years. For comparison, column (1) of Table B.1 shows the same statistics for the universe of all registered firms. By design, the firms in our experimental sample are small, both in terms of the number of employees and the level of VAT payments. Column (3) of Table B.1 provides statistical information on the secondary experimental sample (i.e., the *audit-threat* treatment arm). Despite some statistically significant differences between the two groups, the firms are broadly comparable in size. The main difference between firms in the two experimental samples is that the audit rates were four percentage points higher in the *audit-threat* sample, which is by design because the IRS selected firms classified as high risk for the *audit-threat* treatment arm, and such firms were more likely to have been targeted for audits in the past.

Column (4) reports the same descriptive statistics for the subsample of firms invited to answer the survey, and column (5) reports the p-value of the mean difference between firms that were invited and firms that were not. While the share of firms that paid VAT and the amount of VAT paid were not statistically different between the two groups of firms (p-values of 0.875 and 0.993 respectively), there were some statistically significant differences for other characteristics. Firms invited to the survey had been registered with the IRS for ten fewer months; they were two percentage points less likely to have been audited in the last three years than firms in the main experimental sample on average (p-values of 0.001 and $<$0.001 respectively). Firms in the survey sample had 1.6 more employees and were, on average, 11.5 percentage points less likely to have filed a comprehensive tax return in 2013 (p-value$<$0.001 for both). There are some additional statistically signficant differences in terms of sectors of operation. Firms in the service sector, as opposed to the retail goods sector, were overrepresented in the group of firms invited to participate in the survey (58.8% vs 49.6%, p-value$<$0.001). Overall, while differences in some characteristics are some statistically significant, most do not seem economically significant.

B.2 Delivery Status

An imbalance in the delivery rate of the letters across treatment arms could be a threat to the validity of our results. Since we used the postal service's certified delivery service, we have credible and complete information on the delivery status of each letter. To rule out the possibility of such an imblanace, Table B.2 reports the delivery status of the letters by treatment arm. Columns (1) through (4) report the distribution of delivery statuses by treatment arm for firms in the main sample, while column (5) reports the p-value of the joint equality test. In general, there are no differences by type of letter between the three treatment arms and the *baseline* letter. Only 19.9% of the letters sent to firms in the main sample were returned–a rate that varies little across treatment arms (p-value=0.290). When a letter's delivery status is "Returned" in the postal service dataset, the firm to which it was addressed is excluded from the analysis.

The only statistically significant difference between the treatment arms is the percentage of missing letters; it is smaller for firms that were selected for the *public-goods* message and the *audit-endogeneity* message that for firms selected for the *audit-statistics* message and the *baseline* message. Though statistically significant at a 5% level, these differences are not economically significant (6.7% and 6.2% for *public-goods* and *audit-endogeneity* respectively vs. 7.9% and 7.5% for *audit-statistics* and *baseline*).

Columns (6) through (8) provide an analogous breakdown for the secondary sample, where each delivery status is balanced across treatment arms. Compared to the main sample, the percentage of letters returned to the sender was higher (11.9% vs. 19.9%), perhaps because this sample was specifically selected by the IRS. However, there are no differences between the two treatment arms within the secondary sample.

B.3 Summary Statistics of Tax Payments

Table B.3 reports some descriptive statistics for the pre- and post-treatment periods on the outcome variables used throughout the paper for firms that received the *baseline* letter. The pre-treatment period covers the year immediately before the treatment (October 2014–September 2015) and the post-treatment period covers the twelve months immediately 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 USD 7,770 and the median around USD 4,830, with a standard deviation of USD 8,210. In the subsequent year, the average amount of VAT paid was USD 6,470, the median USD 3,740, and the standard deviation USD 7,770. This represents a reduction of 16.7% in the average VAT payments between pre- and post-treatment periods. Since the group of firms analyzed consists mostly of small and medium-sized firms, this difference could be explained by a high turnover rate.

The bulk of total VAT payments is concurrent payments (about 95% of the total). Retroactive VAT payments represent only about 5% of total VAT payments made by the firms that received the *baseline* letter. The small share of retroactive VAT payments shows that most taxpayers do not make this type of payment. Indeed, the 75th percentile of this distribution is zero. The average amount of retroactive payments made by these firms during the pre-treatment period was USD 400, while during the post-treatment period it was about USD 300. The trends for both concurrent and retroactive payments are consistent with the trend observed for overall VAT payments–and the same holds true of other tax payments and amounts reported in VAT tax returns. Firms paid an average of USD 4,050 of other taxes in the pre-treatment period, and USD 3,300 in the post-treatment period. These amounts include other sales taxes, corporate income taxes, and others. The standard deviation was USD 8,540 in the pre-treatment period and USD 5,430 in the post-treatment period.

VAT tax returns provide additional information that firms report to the IRS regarding total VAT liabilities. The average final VAT liability for the *baseline* group is USD 7,790 in the pre-treatment period and USD 7,190 in the post-treatment period, and its evolution is similar to the sum of monthly VAT payments. The median VAT liability is USD 5,200 in the pre-treatment period and USD 3,840 in the post-treatment period, and the standard deviation is USD 7,910 and 9,930 respectively.

In addition to variables that capture the magnitude of VAT payments–variables more closely associated with intensive margin responses–Table B.3 reports some descriptive statistics about whether firms actually make payments and the number of payments they make. By construction, all the firms in the sample made at least one VAT payment in the pre-treatment period. Ninety-eight percent of the firms made more than three payments, and 89% made more than six payments. In the post-treatment period, 96% of the firms made at least one payment, 91% more than three, and 71% more than six. This pattern is also consistent with the evolution of the number of payments; it also suggests a high turnover rate for the firms in our sample.

This descriptive analysis indicates the importance of VAT to the Uruguayan tax structure. VAT payments are almost twice as high as other tax payments; they represent over 60% of total payments made by the firms in our sample.

B.4 Robustness Checks and Additional Results from Online Survey

B.4.1 Survey Results: Selection into the Survey

In this section, we report a series of additional results and robustness tests for the analysis conducted with the survey data.

Selective responsiveness is a potential concern about the survey data–a threat if the treatment itself induced differential response rates to the survey, or if particular groups of firms were more likely to participate in the survey. We present three pieces of evidence that contribute to our interpretation of the results.

First, we focus on the response rates to the survey and to the two questions about audit probability and penalty rate, that is, the ones relevant to our analysis. Table B.4 reports a series of statistics that shows the selection process from the experimental sample to the final subsample of firms used in the survey analysis, categorized by treatment arm. Invitations with the link to the survey were sent to all firms in our main and secondary samples that had reported their email address to the IRS. The total number of firms invited to the survey was 3,867, or about 23% of our main experimental sample.[75] The share of firms that started the survey (i.e., that answered at least the first two questions in the survey) was 24.5%, and those responses were balanced across treatments. The vast majority of the responses were from individuals who either identified themselves as owners or did not reply to the question about their role at the firm (about 76.5%). An individual who accessed the survey does not necessarily reach its relevant questions, even though they were placed at the beginning: 22.3% of the owners who started the survey did not answer the questions that collected information about prior beliefs of audit probability and penalty rate, which is comparable to the overall non-response rate for the survey. Furthermore, even if respondents reached that point in the survey (i.e., they reported a non-missing value in the previous question), 6.6% skipped the audit probability question and 8.6% the penalty question, a comparable average rate of skipping other questions in the survey (6.1%). We use these answers in our analysis, and in all cases the response rates, however measured, are balanced across treatment arms. We perform a more direct test of selective responsiveness by testing the effect of the treatment on the response rates to the two questions key to our analysis, that is, the ones on audit probability and penalty rate. To do this, we define a dummy variable that takes the value of one if the individual answered the audit probability or the penalty rate question and zero if they did not. We regress this dummy on an *audit-statistics* treatment indicator using

[75] While we sent some invitations to firms in the secondary sample, their responses are not used in our analysis because there were too few firms that satisfied both criteria.

our pooled control group as the reference group (*baseline* or *public-goods* letter recipients). Table B.5 reports the results of these estimates. The *audit-statistics* treatment does not affect the response rate to the audit probability question or the penalty rate question. To address possible differential drop-out rates, we also report the results restricting the analysis to individuals who answered all the survey questions up to the audit probability and penalty rate questions. The results remain the same. Both pieces of evidence suggest that there is no sign of differential behavior in answering the survey by treatment status.

Table B.6 complements this analysis with a balance test for some of the characteristics of the individual survey respondents and their firms who answered the audit probability and penalty rate questions. Column (1) presents information about the age, gender, and city of the respondent as well as some firm characteristics such as size, sector of operation, number of years registered with the IRS, number of locations, and number of employees for firms that received the *audit-statistics* letter. Column (2) does the same for firms that received the *public-goods* letter, column (3) for firms that received the *audit-endogeneity* letter, and column (4) for firms that received the *baseline* letter. Column (5) reports the p-value of the mean test for the four groups. All self-reported characteristics are balanced across treatment arms except for the age of the respondent, the number of years the firm has been registered with the IRS, and the percentage of firms with one employee. For those characteristics, the differences are economically irrelevant, though statistically significant at a 5% level.

Third, Figure B.1 reports a series of placebo tests where we replicate our estimation strategy to measure the effect of the *audit-statistics* message on questions that we do not expect to be affected by our treatment. Reported in other survey modules, these questions aimed to collect information on tax-compliance costs as explained in Section 2. We report the effects of the *audit-statistics* message on the answers to five placebo questions: 1) "On a scale from 1 to 5 where 1 is "Strongly disagree" and 5 is "Strongly agree," to what extent do you agree with the following statement: *Tax compliance generates non-pecuniary costs*"; 2) "On a scale from 1 to 5 where 1 is "Not at all stressful " and 5 is "Very stressful," please rate the level of stress generated by all the steps required to fulfill your tax obligations"; 3) "How much time do you spend informing yourself about your tax obligations?"; 4) "How much time do you spend registering all the transactions made by your firm?"; and 5) "What is the estimated monthly cost of all activities and supplies related to tax compliance?" The estimation strategy is identical to the one used for Figure 5 where the comparison group consists of firms that received the *baseline* letter and firms that received the *public-goods* letter. Figure B.2 contains a summary of the placebo tests compared to the results for the key variables of interest for our analysis (i.e., the ones reported in Figure 5) where, for comparison purposes, all estimates are expressed in standard deviations of the outcome variable for the control group. Figures B.1 and B.2 show that, as expected, the *audit-statistics* message had no effect on the responses to these placebo questions.

B.4.2 Survey Results: Robustness Checks

To analyze the survey, we pooled respondents from the *baseline* and the *public-goods* treatments to form a sufficiently large comparison group. The rationale was that neither of the two messages included information on audit probabilities or tax evasion penalty rates. In this appendix, we assess the robustness of the survey results in Section 5.1 to alternative definitions of the sample and the comparison group.

Panels (a) and (b) of Figure B.3 replicate the results in panels (a) and (b) of Figure 5. The shaded gray bars show the distribution of perceptions for the sixty-nine survey respondents that received the *baseline* letter only (Figure 5 relied on the 137 observations from the pooled *baseline* and *public-goods* groups). The dashed curve red line corresponds to the distribution of signals sent to the firms in the *audit-statistics* letters. Although slightly smaller than in the pooled control group, the average perceived audit probability for the *baseline* letter group (37.7%, in panel (a)) is still substantially higher than the 11.7% for the overall sample, a statistically significant difference. The results are also consistent with the main results when we look at the perceived penalty rate in panel (b). There are no statistically significant differences between the perceived penalty rate by the firms that received the *baseline* letter and our estimates from the overall data.

Second, we perform an additional robustness test of the results in Figure 5. To increase the likelihood that survey respondents were the ones who received our experimental messages, we replicate the analysis, this time restricting our sample to survey respondents who, in the survey, self-identified as firm owners. Presented in panels (c) and (d) of Figure B.3, the results with this restricted sample, which reduced the treatment group from 365 to 341 observations and the pooled control group from 137 to 125, are very similar to the results reported in the body of this paper. Our *audit-statistics* treatment significantly reduced the perceived average probability of audits, although it did not affect the average perception of penalty rates. Finally, panels (e) and (f) of Figure B.3 report the results of a similar analysis, but this time using all answers regardless of the respondent's self-reported role at the firm, that is, not only owners, but also managers, inhouse accountants, external tax advisors, and other employees. This modification increases the size of the treatment group from 365 to 465 and the size of the control group from 137 to 179. The results in terms of magnitude, direction, and statistical significance are the same.

An additional concern is that respondents may not have taken the survey seriously. One way of testing this is to analyze the robustness of our results without the respondents who indicate 50% as their perceived audit probability or penalty rate (Bruine de Bruin et al., 2002; Bruine de Bruin and Carman, 2012).[76] To address this concern we take two steps. First, we test whether the degree of certainty for individuals who indicate a 50% perceived audit probability or penalty rate is different from the degree of certainty for those who provide non-50% responses.[77] We present the results of this test in Figure B.4. Panel (a) of Figure B.4 reports the distributions of the degree of certainty in the answers provided to the audit probability question for 50% responses and non-50% responses. The average certainty for people with non-50% answers was 2.37 compared to a mean of 2.03 for people who answered 50% (measured on a scale from one to four, where one corresponds to an answer of "Not at all sure" and four to an answer of "Very sure"). This difference is statistically significant (p-value<0.001) and suggests that 50% answers indeed corresponded to individuals who did not feel sure about the answer they provided. The results are similar for Panel (b) of Figure B.4, which reports the results for certainty about the reported perceived penalty rate. Individuals with a 50% response feel less certain of their answers than individuals with a non-50% response. That does not necessarily mean, however, that individuals who answered 50% did not take the survey seriously. It simply reflects that they felt less certain about their answers. To rule out the possibility that our results could be driven by this type of response, Figure B.5 replicates our analysis reported in Figure 5 without individuals who provided 50% responses. Our findings hold, although our estimates are less precise due to the reduction in the number of answers included in the analysis.

If firms were rational, all these results would imply that firms would have paid less tax as a consequence of altering their beliefs. This is not, however, what we observe, which supports the hypothesis that the results are driven by the fear channel rather than by a rational re-optimization.

B.4.3 Survey Results: Beliefs About Audit Endogeneity

As in the case of the *audit-statistics* treatment arm, we conducted a survey of letter recipients that 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 its evasion amount, the effect on its probability of being audited would..." The possible answers were: Increase significantly; Increase slightly; Not change;

[76] For simplicity's sake, in the remainder of this section we refer to such cases as 50% answers or 50% responses.

[77] We directly elicited the degree of certainty about perceived audit probability (in Q3) and penalty rate (in Q5) in the survey. These questions are reported in the survey questionnaire in Appendix A.7.

Diminish slightly; Diminish significantly.

The distribution of responses to this question is depicted in Figure B.6. The distribution of perceptions in the pooled control group–firms that received the *baseline* and *public goods* letters–suggests that firms were already aware of this endogeneity. In other words, there are no statistically significant differences in the distribution of perceptions for the *audit-endogeneity* group and the pooled control group group. On a scale from one to five, where one indicates that more evasion significantly increases the probability of being audited and five indicates that more evasion significantly diminishes the probability of being audited, the average belief was 1.45 in the pooled control group and 1.41 in the *audit-endogeneity* group (p-value of the difference=0.67).

B.4.4 Survey Results: Relation Between Signal and Self-Reported Perception

Figure B.7 shows the raw relation between the signal in the letter and the perceptions of audit probability and penalty rate reported in the survey. The x-axis depicts the signal in the letter and the y-axis the self-reported parameter. Panels (a) and (b) of Figure B.7 report the raw scatterplots for the audit probability and penalty rate respectively, where each dot represents a pair of values and the size of the dot is proportional to the number of individuals who received a given signal and their self-reported perception. While there is a positive correlation between the signal and the reported perception for the audit-probability parameter (depicted in panel (a)), individuals clearly overestimate the chances of being audited even after receiving the letter. Panel (b) also reports a positive corelation between the signal in the letter and the self-reported perception of penalty rate. In contrast to perceived audit probabilities, perceived penalty rates seem less disperse and closer to the actual value (about 30%). Panels (c) and (d) of Figure B.7 provide binned scatterplots that depict the results of regressing the self-reported perception of the parameter onto the signal in the letter for audit probability and penalty rate respectively. For the audit probability (depicted in panel (c)), individuals seem to have an adjustment rate of 40%. That means that for each additional percentage-point increase in the signal, individuals reported 0.4 additional percentage points in their perceived audit probability. The p-value associated with this coefficient does not allow us to reject the null hypothesis of no updating (p-value=0.169). For the penalty rate, the results depicted in panel (d) suggest that the information in the letter did not affect individuals' perceptions at all: for each additional percentage-point increase in the penalty rate as reported in the letters, individuals adjusted their perceived penalty rate by less than 10%. With a standard error of 0.20, the null hypothesis of the signal having no effect on perceptions cannot be rejected (p-value=0.965).

The information in the letter may also have altered individuals' degree of certainty about their perceptions of audit probability and penalty rates. Figure B.8 shows the effect of the *audit-statistics* letter on answers to Q3 and Q5, which ask directly how certain individuals are about their responses to the questions about perceived audit probability and penalty rate.[78] There are four possible answers for each question: "Not sure at all," "Somewhat sure," "Sure," and "Very sure." Figure B.8 suggests that, if anything, the information provided in the letter made firms less sure of their perceptions of p but did not affect their confidence about their perceived θ (p-values of 0.0688 and 0.8120 respectively).

B.5 Dynamics of the Effect-Raw Comparison Treatment vs Control

Figure B.9 plots the evolution of total VAT payments by treatment status, grouped by pairs of months. This raw visualization of the data allows for a very simple comparison of the treatment and comparison groups at each pair of months depicted. We group VAT payments by pairs of months because many firms in our sample are only required to make VAT payments on a bimonthly basis. Period zero, represented by the vertical dashed line, is defined as August–September 2015, that is, the period when the letters were delivered by the post office. In all panels, the lighter solid line represents the evolution of VAT payments for the comparison group, while the darker dashed line represents the evolution of VAT payments for the treated group. Panels (a) through (c) of Figure B.9 depict the treatment-control comparison in the main sample (i.e., *audit-statistics* vs. *baseline*, *audit-endogeneity* vs. *baseline,* and *public-goods* vs. *baseline* respectively). Panel (d) of Figure B.9 provides a comparison of the two treatment groups in the secondary sample: 50% audit probability threat vs. 25% audit probability threat. In all cases, we exclude firms whose letters were marked as "Returned" by the post office.

The results in Panels (a) through (c) of Figure B.9 indicate no difference in VAT payments between firms that received any of the treatment messages and firms that received the *baseline* letter before the letters were sent (this is formally contrasted by means of the balance tests reported in Table 1 and the falsification tests reported in Table 2). Immediately after receiving a letter, however, a wedge in VAT payments shows up between the two groups, and firms that were treated with any of the messages start to pay larger amounts of VAT compared to firms that received the *baseline* letter. These differences seem to be larger in the first twelve months after the experiment, although a smaller gap persists after the first post-treatment year. These patterns are consistent with the ones that can be derived

[78]These questions are found in the survey questionnaire in Appendix A.7.

from the difference-in-differences specification reported in Table 2. Panel (d) of Figure B.9 focuses on the secondary sample. The figure suggests that the evolution of VAT payments by treatment arm in this special sample is noisier than in the main sample; no clear pattern can be observed immediately after the letters were sent or in subsequent periods. This is consistent with the results reported in Table 3, where the results clearly indicate that the *audit-threat* letter did not have a differential effect on the group of firms that received the higher signal of probability of being audited.

B.6 Robustness Checks: Regression Analysis

B.6.1 Robustness Checks of Main Specification

To assess the robustness of the results for the main specification in Section 4, Table B.7 presents alternative estimates for the effects of the *audit-statistics*, *audit-endogeneity,* and *public-goods* treatments based on different specifications. The first column presents estimates of the treatment effects based only on the extensive margin of VAT payments: i.e., the outcome is coded as one if the firm made at least one payment in the post-treatment period, and zero otherwise. 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 that had not made at least three payments in the twelve months before the treatment assignment. To complement these results, columns (2) and (3) present the average treatment effects on alternative outcomes related to the number of payments made by the firms. Column (2) reports the average treatment effect on the probability of making at least three payments, and column (3) on the probability of making at least six payments in the same period.[79] The effects of the three different messages on the extensive margin and on the number of payments are close to zero and statistically insignificant.

The specifications in columns (4), (5), and (6) of Table B.7 use the number of VAT payments as the dependent variable. Column (4) corresponds to our main Poisson specification. Column (5) presents estimates based on OLS regressions, and column (6) presents estimates based on Tobit regressions. The Poisson model has a key advantage in this context: it deals naturally with the bunching of payments at exactly zero while allowing for the effects to be proportional. By contrast, the OLS specification does not deal with bunching at zero and does not allow for the effects on number of payments to be proportional. Because of the

[79]We do not include an outcome variable that reflects the treatment effect on the probability of making twelve payments during the first post-treatment year because, as mentioned, many firms in our sample are only required to make VAT payments on a bimonthly basis (only 25% of the sample makes payments every month). See Table B.3 for more details about the distribution of monthly payments.

nature of our sample, many firms are required to make VAT payments on a bimonthly basis. The Tobit specification is more appropriate than OLS since it takes into account the censored nature of the data at zero but does not allow for the effects to be proportional .

Columns (4), (5), and (6) of Table B.7 present estimates based on these specifications. They show the results to be identical in terms of sign and statistical significance of the coefficients, indicating robustness to these three alternative specifications. If anything, the effects are more statistically 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 magnitude, they are roughly consistent. For instance, the Tobit model suggests that the *audit-statistics* message has an effect of USD 451 (p-value=0.003). Since the average outcome is USD 6,470, this Tobit coefficient amounts to an effect of about 6.9%, which is in the same order of magnitude as the Poisson model (it indicates an effect of 7.0% for the *audit-statistics* message (p-value=0.001)).

Finally, column (7) of Table B.7 reports the results of estimating our model on the final VAT liability calculated from an alternative administrative data source: annual tax returns. This outcome overcomes the possible concern that tax delinquency could be a driver of our results on the effect of retroactive payments specifically. The time frame for this outcome is completely different from the one for the monthly VAT payments used in our main specification, and hence the amounts (and thus the results) do not necessarily match. However, the results in column (7) of panel (a) of Table B.7 indicate that the effect of the *audit-statistics* message is 5.6% with this alternative measure of the outcome variable, which is similar in sign and magnitude to our main result. This effect is precisely estimated and statistically significant at the 5% level. The effect of the treatment in the falsification test is indistinguishable from zero.

The results in panel (b) of Table B.7 indicate that the effect of the *audit-endogeneity* message on the VAT reported in the annual tax return is similar to its effect on the VAT payments. The point estimate of the coefficient is 5.9%, which is smaller than the result for our main specification. In this case, however, the coefficient is less precisely estimated and statistically indistinguishable from zero at conventional levels. Finally, the results in panel (c) of Table B.7 indicate that the effect of the *public-goods* message on the annual VAT liability reported in the tax return is close to zero, a finding consistent with small effects observed immediately after receiving the letter that fade a few months later.

B.6.2 Alternative Specifications for the Effects of Signals on Audit Probabilities and Penalty Rates

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

First, panel (a) in Table B.8 replicates the analysis in Table 3 for the additional specifications we used in Table B.7. The results are essentially the same as the ones reported in Section 5: i.e., the specific information on audit probabilities and penalty rates in the *audit-statistics* message had no effect on compliance. At both the extensive and the intensive margin, all coefficients associated with the treatment variable, regardless of the specification used or the source of the dependent variable, are statistically insignificant. This is also true for the effects of the *audit-threat* message (panel (b) of Table B.8). Conditional on being treated, the information in the letter does not affect firms' compliance behavior.

Second, Table B.9 presents the results for an alternative specification of the elasticity estimation in Table 3. Instead of estimating the elasticities with respect to p and θ separately, we estimate the elasticity with respect to the product (expected penalty) $p * \theta$ in a regression of the form:

$$Y_i = \alpha + \tau_{p \cdot \theta} \cdot p_i \cdot \theta_i + \delta \cdot Post_t + \gamma_{p \cdot \theta} \left(p_i \cdot \theta_i \cdot Post_t \right) + \sum_{g=1}^{5} I_{\{i \in g\}} \left(\pi_g + \kappa_g \cdot Post_t \right) + \epsilon_i \quad \text{(B.1)}$$

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

Panel (a) of Figure B.10 provides additional evidence on the effect of the *audit-statistics* message (relative to the *baseline* letter). In the spirit of Figure 4, Figure B.10 summarizes the raw data, this time dividing the sample into two groups: firms that received letters with low signals of p ($p \leq$11.7%) and firms that received high signals of p ($p >$11.7%). Note that the evolution of the outcome variable is extremely similar for the two groups.

Panel (b) of Figure B.10 presents the equivalent analysis for firms that received messages with low and high values of θ (below and above the mean penalty rate in our sample, 30.6%). Again, the effects are very similar for the two groups.

The results in Table B.8, Table B.9, and Figure B.10 provide further evidence that what is at work is the fear channel rather than a rational re-optimization as a consequence of exposure to some information on tax enforcement mechanisms.

B.7 Additional Test: Breakdown of Other Taxes

In this section, we report additional and more detailed evidence about the treatment effects by analyzing the effects of the letters on different types of payments firms made regularly to the IRS. Regular payments by the firms to the IRS are comprised of VAT payments, corporate income tax payments, property tax payments, and personal income tax withholdings on behalf of employees. Table B.10 replicates the difference-in-differences estimates of the average effect of the different treatment arms presented in Table 2 but for different aggregations of payments for the first year after the experiment. The table reports both the treatment effect and the estimates from the falsification test comparing two pre-treatment periods, as does Table B.11 but for the results presented in Table 3. Column (1) reports the average effects on the total amount of tax paid by the firms, that is, the sum of VAT, corporate income tax, property tax, and personal income tax withholding payments. Columns (2) and (3) separate total payments into VAT and other taxes, while columns (4), (5), and (6) break down other taxes into their three components.

The results provided in Table B.10 show that the treatment effects caused by the *audit-statistics*, *audit-endogeneity,* and *public-goods* letters are not restricted to VAT payments; payments of other taxes increased as well, indeed by a slightly larger magnitude. When the "other taxes" variable is broken down into its components, the effects are too imprecise to identify any statistically significant effect on any single tax. Overall, all the coefficients are positive and of the right magnitude with no pre-treatment imbalances, but the effect of the *audit-statistics* message on corporate income tax is the only one that is statistically significant. The interpretation of the results reported in Table B.11 is similar, indicating that the information in the letters did not seem to have an effect on any tax payment except for personal income tax withholding. Though that effect is statistically significant in some specifications, there is no clear and consistent pattern across specifications.

B.8 Additional Test: Heterogeneity with Respect to Firm's Characteristics

To provide a more detailed description of how the messages affected different types of firms, we report here the treatment effects disaggregated by different firm characteristics. Because administrative records contain only a few observable firm characteristics, our analysis is based on four characteristics: number of employees, number of years the firms has been registered with the IRS, type of tax return filed, and activity sector. For number of employees and years registered with the IRS, we split the sample in two according to whether the observed value falls above or below the median (2.5 and 12.7 respectively). The IRS has different VAT schedules depending on firm size, and that is reflected in the type of tax return firms are required to file. Even for the small and medium-sized firms considered in this experiment, there are different types of tax returns. For a heterogeneity analysis, we divide firms into two groups: firms required to file a more comprehensive tax form (mostly larger firms) and firms required to file a simplified tax form according to special VAT regimes such as fixed VAT and professional independent VAT that have a separate form. As reported in Table 1, about 68% of firms filed a comprehensive VAT tax return in 2013. Finally, we also have information about the sector in which the firm operates. We split the sample into three groups: non-retail firms that sell goods (about 29%), retail firms (about 22%), and service providers (49%).

Figure B.11 depicts the average effect of the letters on VAT payments in the first year after the experiment (and the associated 95% confidence intervals) for each of the groups as defined above (similar to the "Post-Treatment" coefficient in column (1) of Table 2). Figure B.12 reports the same estimates but for the falsification test that compares treated and control firms in two pre-treatment periods (similar to the "Pre-Treatment" coefficient in column (1) of Table 2). Figure B.13 reports the estimated elasticities with respect to the audit probability and penalty rate in the letter for the same groups of firms, which is analogous to the results reported in Table 3. Figure B.14 reports the falsification test for the estimated elasticities. Figures B.11 and B.12 are based on four regressions. Each regression is an augmented version of equation (1) altered to include an additional interaction between the coefficient of interest and a dummy for the group of interest and for all the corresponding interaction terms. Figures B.13 and B.14 are similar, but based on equation (2) instead. This is our baseline specification for the elasticities of VAT payments to the parameters provided in the letters. Panel (a) of Figure B.11 compares the *audit-statistics* message with the *baseline* letter, while results in panels (b) and (c) replicate that comparison for *audit-endogeneity* and *public-goods* messages respectively. The notes in the figure report the p-value of the test that compares whether the treatment effects are different across groups within a single category.

The coefficients depicted in Figure B.11 show that while the letters have a positive effect on VAT payments overall, that effect is, it appears, homogeneous across firm types. Out of fifteen tests of equality of coefficients, only one is rejected at a 10% level (firms filing a more comprehensive tax return seem to respond slightly more than firms that do not file comprehensive tax returns). Figure B.13 is also consistent with homogeneous responses across groups of firms. Estimated elasticities are close to zero on average, as our main specification suggests, and there are no differences between groups based on any of the variables considered, except for firms in the service sector for which the test of equality of coefficients can be rejected at a 1% level (p-value=0.013). Falsification tests reported in Figures B.12 and B.14 are reassuring since they show that there are no differences in the "fake" treatment effect between firms in the pre-treatment period.[80] The general conclusion is, then, that all firms respond in a fairly similar way, although we do not have enough power to capture the differences that do exist.

B.9 Additional Test: Heterogeneity with Respect to Prior Beliefs

B.9.1 Measuring Prior Beliefs

In the *A&S* framework, firms with different prior beliefs about the probability of being audited should react differently to signals and information on this probability. To test this hypothesis, we need a measure for the prior beliefs of a particular firm. We construct a proxy of prior beliefs based on the firm's own audit history.

The intuition behind this approach is that, since there is little publicly available information on audit probabilities, firms likely form their beliefs based on their own audit experience. For instance, when a firm registers with the tax authority, its initial belief may follow the beta distribution with parameters $\{\alpha_0, \beta_0\}$. Assume that firm i has been registered for T_i years before our mailing campaign, and during this period it has experienced $N_i \leq T_i$ audits. If firm i is Bayesian, its belief about annual probability of being audited should follow a beta distribution with parameters $\{\alpha_1 = \alpha_0 + N_i,\ \beta_1 = \beta_0 + T_i - N_i\}$. The mean of that belief should be $\frac{\alpha_0+N_i}{\alpha_0+\beta_0+T_i}$. This implies a belief about the probability of being audited at least once in the following three years of $\hat{p}_i = 1 - \left(1 - \frac{\alpha_0+N_i}{\alpha_0+\beta_0+T_i}\right)^3$. In our main specification, we generate these proxies by setting $\{\alpha_0 = 0.13, \beta_0 = 1\}$. This baseline calibration generates an average belief that matches the actual average probability according to our administrative data, and we offer robustness tests using alternative calibrations.[81]

We can use the survey data to validate this proxy for prior belief. Among the 145 responses from the pooled control group, 10.3% of firms reported that they had been audited in the past, and these firms reported a higher average perceived probability of being audited at least once in the following three years (63.9%) relative to firms that had not been audited recently (38.1%), a large and statistically significant difference (p-value<0.001).[82] This ev-

[80]While in general the results from the falsification tests are reassuring and centered around zero, there is a common pattern of statistically significant "false" effects when the sample is split by type of tax return filed. This is due to an anomalous change in VAT payments in the *baseline* group for firms that did not file a comprehensive tax return. VAT payments for the *baseline* group that did not file a tax return fell significantly between the two years considered for the falsification test, which is why we observe a positive "false" treatment effect. This is observed across panels because the baseline group is used as the comparison group in panels (a) through (c).

[81]Since we only have information about audits for firms in our sample for the previous fifteen years, we set the maximum number of years registered with the IRS at fifteen.

[82]As reported in a follow-up paper (Bergolo et al., 2018), the indicator of recent audits is the single most important predictor of perceived audit probabilities among a host of different factors. Recent audit

idence suggests that, consistent with our proxy, firms are using their own audit history to form beliefs about the probability of being audited in the future.

B.9.2 Results

In the *A&S* framework, the effect of the *audit-statistics* letter on tax compliance should be larger for firms with relatively low priors for the audit probability ($\hat{p}$) compared to those with relatively high values. Since the *audit-statistics* message conveyed a signal of $\hat{p} = 11.7\%$ on average, its effect on compliance should have been greater for firms with $\hat{p} < 11.7\%$, since the signal should, on average, increase their perceived audit probability. The effect of the *audit-statistics* message should be negative for firms with $\hat{p}>11.7\%$: on average, the signal should reduce their perceived probability of being audited.

Figure B.15 presents the results from this exercise. Panel (a) of Figure B.15 shows a binned scatterplot of the treatment effect of the *audit-statistics* letter for the four quartiles of $\hat{p}$.[83] This figure includes a vertical dashed line at 11.7%, the average audit probability conveyed in our letters. In contrast to the predictions from *A&S*, we fail to find a negative relationship between the effect of the *audit-statistics* message and the value of the prior belief: the slope is negative (-0.0006), but economically small and statistically insignificant (p-value of 0.525). Panel (b) of Figure B.15 presents a more direct test by combining the heterogeneity in prior beliefs with the heterogeneity in signals. Instead of grouping firms by prior belief, we group them by the difference between that prior belief and the specific signal sent to each firm in the personalized letter. The intuition is that the difference between the prior belief and the signal is the "surprise" conveyed by our information treatment. The figure includes a vertical dashed line at zero, the point where firms receive signals that are equal to their prior beliefs. The effect of the *audit-statistics* letter on compliance should be decreasing in $\hat{p}-p_{signal}$, positive for the group with $\hat{p}-p_{signal} < 0$ (i.e., those for whom the signal was higher than the prior belief), and negative for the group with $\hat{p} - p_{signal} > 0$ (i.e., those receiving a signal indicating that they were overestimating audit probability). The results in panel (b) are consistent with the results from panel (a): the slope of the relationship (indicated by the dashed red line) is zero (p-value=0.986).

As a robustness check, we provide an alternative calibration of the Bayesian model. In the above results, we selected values for the parameters α_0 and β_0 to "center" beliefs around

experience also has a positive effect on perceived penalty rate, but this effect is less significant: respondents from firms that were audited recently report an average perceived penalty rate of 40.0%, compared to 29.4% for respondents from firms that were not audited recently (this difference is statistically insignificant (p-value=0.201)).

[83] We base our analysis on quartiles of the probabilities because there is substantial bunching at zero, which forced us to divide the sample in four to get even-sized groups.

the true audit probability in our sample (11.7%). We present the results from an alternative calibration, which centers the perceived probability around the average value obtained from the control group in our post-treatment survey–an average perceived audit probability of 40.5% for the comparison group. The results are presented in Figure B.16. If anything, this alternative calibration makes *A&S* even less plausible: as shown in panel (b), we would expect almost everyone to adjust perceived probability downward, and thus reduce tax payments–exactly the opposite of what we find.

A more structured analysis is presented in Table B.12. For reference, column (1) of Table B.12 shows the baseline specification: i.e., in the whole sample, the *audit-statistics* message increased tax payments by 7.0%. We then divide the sample into two groups: firms with prior beliefs about the probability of being audited that were lower than the probability reported in the letter they received, and firms with prior beliefs about the probability of being audited that were equal to or higher than the probability reported in the letter they received. Since we divide the groups according to a characteristic of firms that received the *audit-statistics* message, the outcomes for both groups are compared to those of the firms that received the *baseline* letter. According to *A&S*, increasing taxpayers' perceived probability of being audited should result in higher tax payments, and the opposite should hold true when their perceived probability of being audited is reduced. The expected effect of the *audit-statistics* message is, therefore, positive on firms with $\hat{p} < p$ and negative on firms with $\hat{p} \geq p$. Column (2) in Table B.12 shows that the average effect of the *audit-statistics* message on VAT payments in the first year after receiving the letter is 6.1% for taxpayers with relatively low prior beliefs. Column (3) reports that the average effect for taxpayers with relatively high beliefs is 8.6%. The effect on firms with relatively low prior beliefs is positive but lower in magnitude than the effect on firms with relatively high prior beliefs, and the sign of the effect on the latter is the opposite of *A&S*-based predictions. Furthermore, the differences in magnitude are statistically and economically insignificant (p-value=0.347). If these results provide any evidence, it counters the *A&S* predictions.

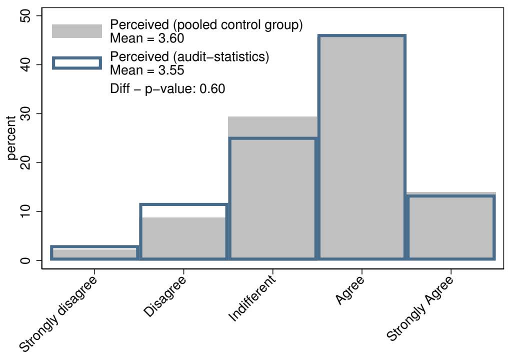
We can reproduce a similar exercise without a Bayesian learning model. We compare the responses of firms with and without prior audit experience. In columns (4) and (5) of Table B.12, we compare the effects of the *audit-statistics* message on firms that were audited in the recent past (23.7% of the sample was audited at some point in the fifteen years before the treatment) and firms that were not audited during that period (the remaining 76.3% of the sample).[84] The intuition is that firms that were audited in the recent past believe the probability of being audited is higher than those that were not audited. The null hypothesis is thus that the effect of our *audit-statistics* messages is stronger for firms that were not

[84]That time frame reflects how far back the available IRS administrative records reach.

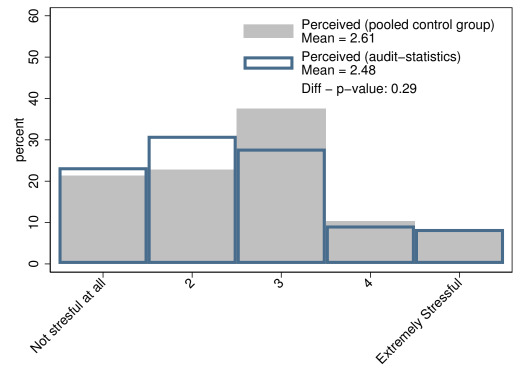
audited, since such firms would have had lower prior beliefs about audit probabilities and thus increase their perceived probabilities in response to the *audit-statistics* messages. We could even expect to see negative effects on compliance for firms that were audited in the past, because they were likely to have had high prior beliefs about audit probabilities that our information treatment should have reduced. The results in Table B.12 indicate that there are no heterogeneous effects with respect to recent audit experience. The difference in treatment effects for the two groups is small and statistically insignificant: the effect is 8.2% for the group of firms that was previously audited, and 6.6% for those with no prior audit experience (p-value of difference=0.729).

Figure B.1: Survey Results: Placebo Variables by Treatment Group

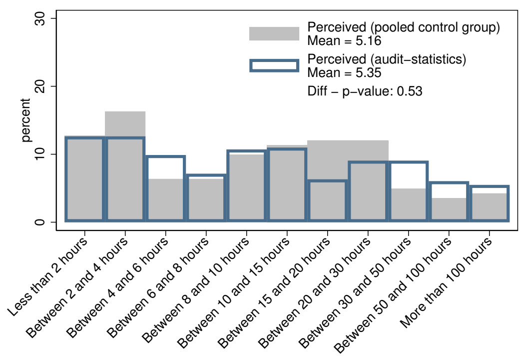
a. Non-Pecuniary Costs



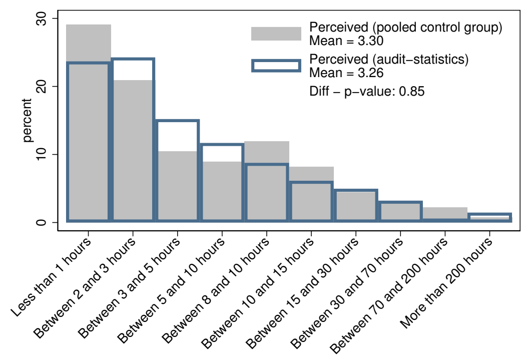
b. Stress



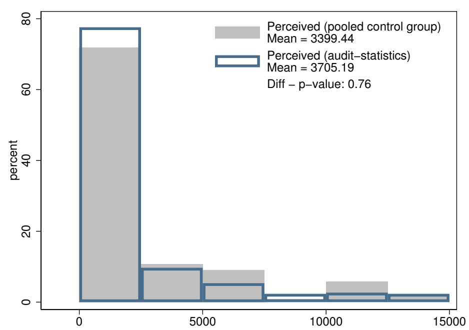
c. Time Required to Inform on Taxes



d. Time Registering Transactions

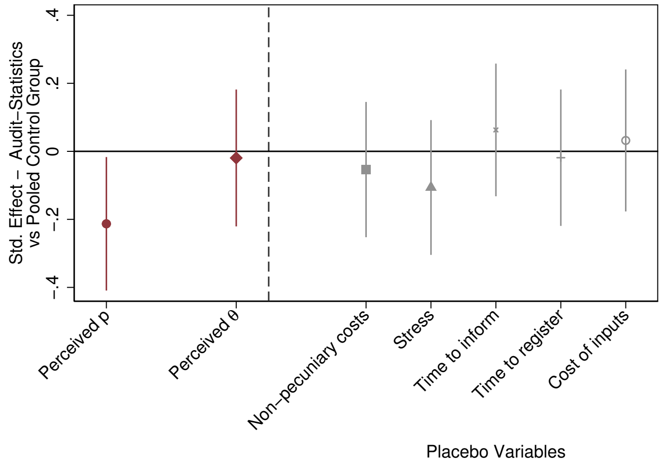


e. Cost of Inputs to File Taxes



Notes: Placebo tests where we estimate the effect of receiving the *audit-statistics* letter on questions that we do not expect to be affected by our treatment. These questions were reported in other modules of the survey and were aimed to collect information about tax compliance costs. Panel (a) reports the effects on the answers to a question that asked: "On a scale from 1 to 5 where 1 is 'Strongly disagree' and 5 is 'Strongly agree,' to what extent do you agree with the following statement: *Tax compliance generates non-pecuniary costs*?" Panel (b) presents the responses to the following question: "On a scale from 1 to 5 where 1 is "Not stressful at all" and 5 is "Very stressful," please rate the level of stress created by all the steps required to fulfill your tax obligations." Panel (c) presents the responses to the following question: "How much time do you spend informing yourself about your tax obligations?" Panel (d) presents the responses to the following question: "How much time do you spend registering all the transactions made by your firm?" Panel (e) presents the responses to the following question: "What is the estimated monthly cost of all the inputs that you use in activities related to tax compliance?" The estimation strategy is exactly the same as the one used for Figure 5 (see corresponding notes for more detail). "Perceived (pooled control group)" refers to survey respondents who received the *baseline* letter or the *public-goods* letter during the experimental stage. "Perceived (*audit-statistics*)" refers to respondents who received *audit-statistics* letters. We report the mean responses and the p-value of the difference between the two groups. Analysis is restricted to individuals who are owners (or who did not answer the ownership question) and who answered the questions on perceived audit probability and penalty rate.

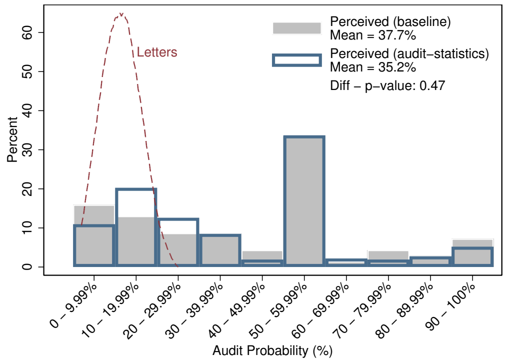
Figure B.2: Survey Results: Placebo Variables, *Audit-Statistics* vs. *Baseline* and *Public-Goods*



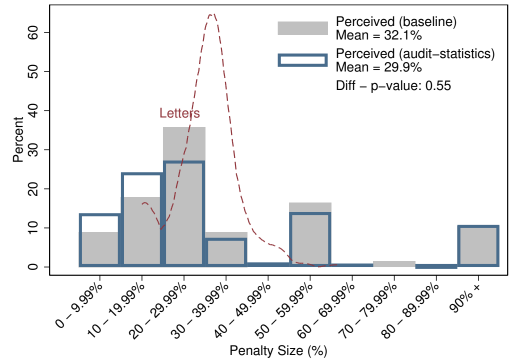
Notes: In this figure, we compare the treatment effects of the *audit-statistics* letter on perceptions of audit probability and penalty rate (coefficients reported to the left of the vertical dashed line) to the placebo outcomes reported in Figure B.1 (see corresponding notes for more details about the placebo outcomes). For all of the outcomes reported in the figure, we used the same estimation strategy as in Figure 5 (see corresponding notes for more details). The comparison group is comprised of respondents who received the *baseline* letter or the *public-goods* letter during the experimental stage. The treatment group is comprised of respondents who received *audit-statistics* letters. For comparison purposes, all outcome variables are expressed in terms of the standard deviations of the control group. In all cases, the analysis is restricted to individuals who are owners (or who did not answer the ownership question) and who answered the two questions that were relevant to our analysis (the ones about perceived audit probability and penalty rate).

Figure B.3: Survey Results: Perceived p and θ, Alternative Samples and Comparison Group

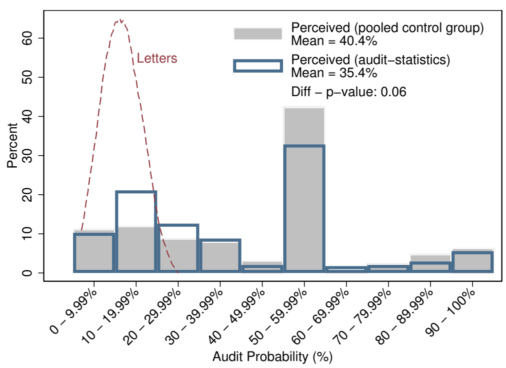
a. *p: Audit-Statistics* vs. *Baseline*



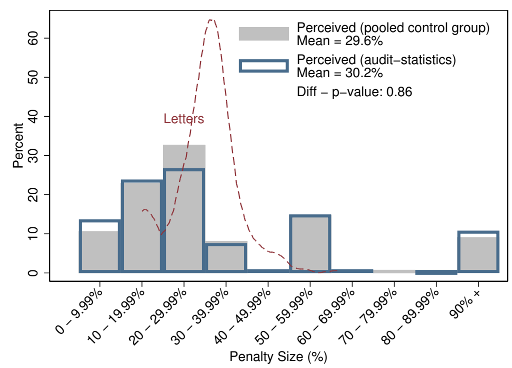
b. θ: *Audit-Statistics* vs. *Baseline*



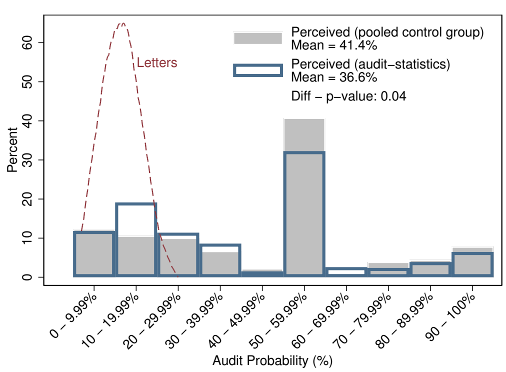
c. *p: Audit-Statistics* vs. *Baseline* and *Public-Goods*, Owners Only



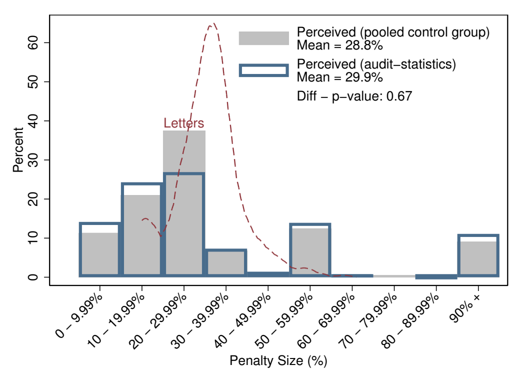
d. θ: *Audit-Statistics* vs. *Baseline* and *Public-Goods*, Owners Only



e. *p: Audit-Statistics* vs. *Baseline* and *Public-Goods*, Full Sample



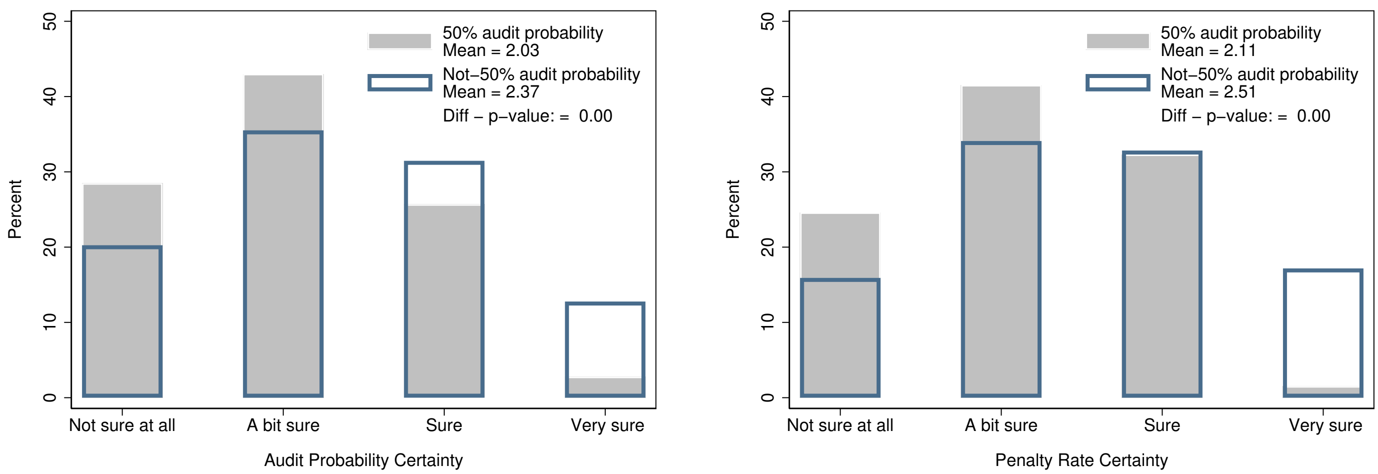
f. θ: *Audit-Statistics* vs. *Baseline* and *Public-Goods*, Full Sample



Notes: In panels (a) and (b), "Perceived (*baseline*)" (N=69) refers to survey respondents who received the *baseline* letter during the experimental stage while "Perceived (*audit-statistics*)" (N=365) refers to respondents who received the *audit-statistics* letter. In panels (c) and (d), we use a pooled control group comprised of recipients of the *baseline* and *public-goods* letters, but we restrict the sample to survey respondents who self-identified as owners (N of *baseline* group = 61, N of *public-goods* group = 64, N of *audit-statistics* group = 341). Panels (e) and (f) use the full sample regardless of the self-reported occupation in the survey, that is, it includes owners, managers, inhouse and external accountants, and other employees (N of *baseline* group = 89, N of *public-goods* group = 90, N of *audit-statistics* group = 465). We also report the mean of the perceptions for each parameter and the p-value of the difference between the groups in each panel. The answers correspond to survey Q2 and Q4 (see full survey questionnaire in Appendix A.7). In panels (a), (c), and (e), the x-axis represents the probability of being audited; in panels (b), (d), and (f), it represents the average penalty rate. The red line represents the density function of the information displayed in the *audit-statistics* letters, measured in the y-axis on the right (hidden for the sake clarity). In all cases, the analysis is restricted to letters that were not returned by the postal service.

Figure B.4: Certainty in Perceptions Reported by Having Answered 50%

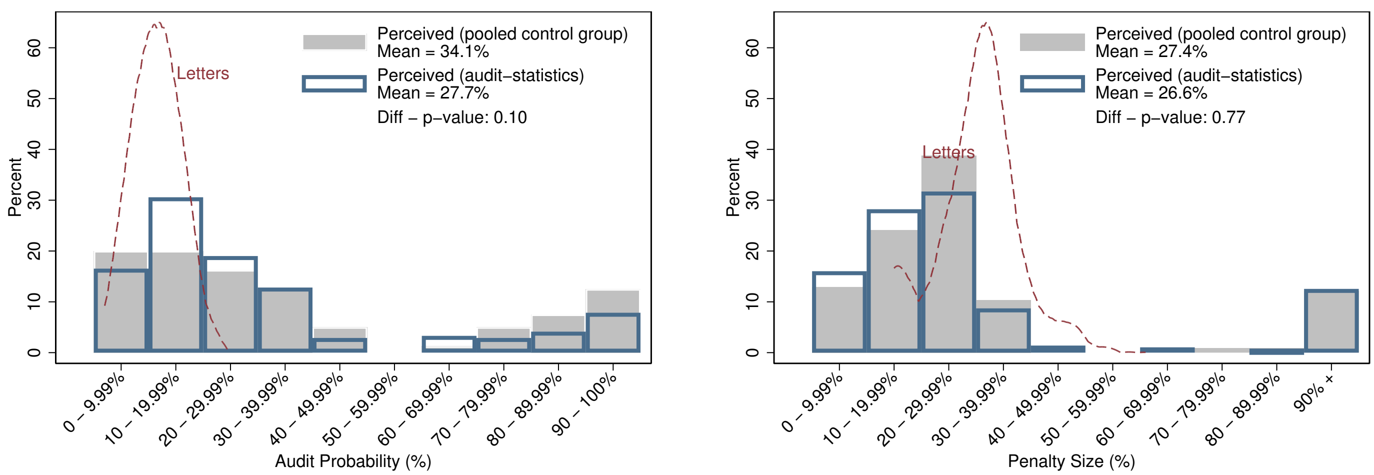
a. Audit Probability (p) **b.** Penalty Rate (θ)



Notes: The histograms present the distributions of certainty about perceived audit probability and penalty rate as elicited in the survey. Panel (a) reports the distribution of answers to survey Q3 (see full survey questionnaire in Appendix A.7), splitting the sample between individuals who reported a perceived audit probability of 50% (N=180) and individuals who reported a different audit probability (N=322). Panel (b) presents a similar breakdown for survey question Q5, that is, it also splits the sample between individuals who reported a perceived penalty rate of 50% (N = 65) and individuals who reported a different perceived penalty rate (N = 397). In both panels, the sample is restricted to individuals who belong to our *audit-statistics* treatment arm or to the pooled control group (i.e., those who received either the *baseline* or the *public-goods* letter). We report the mean responses for the various groups, measured on a four-point scale where one corresponds to "Not at all sure" and four to "Very sure." We also report the p-value of the difference between the two groups.

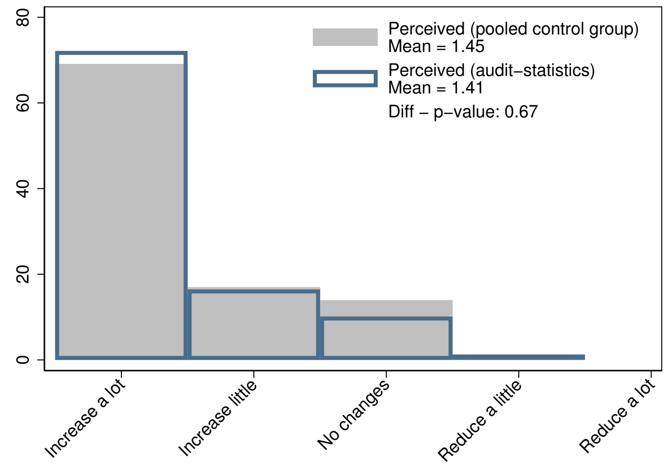
Figure B.5: Survey Results: Perception of Audit Probabilities and of Tax Evasion Penalty Rates by Treatment Group - Excluding 50% Responses

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



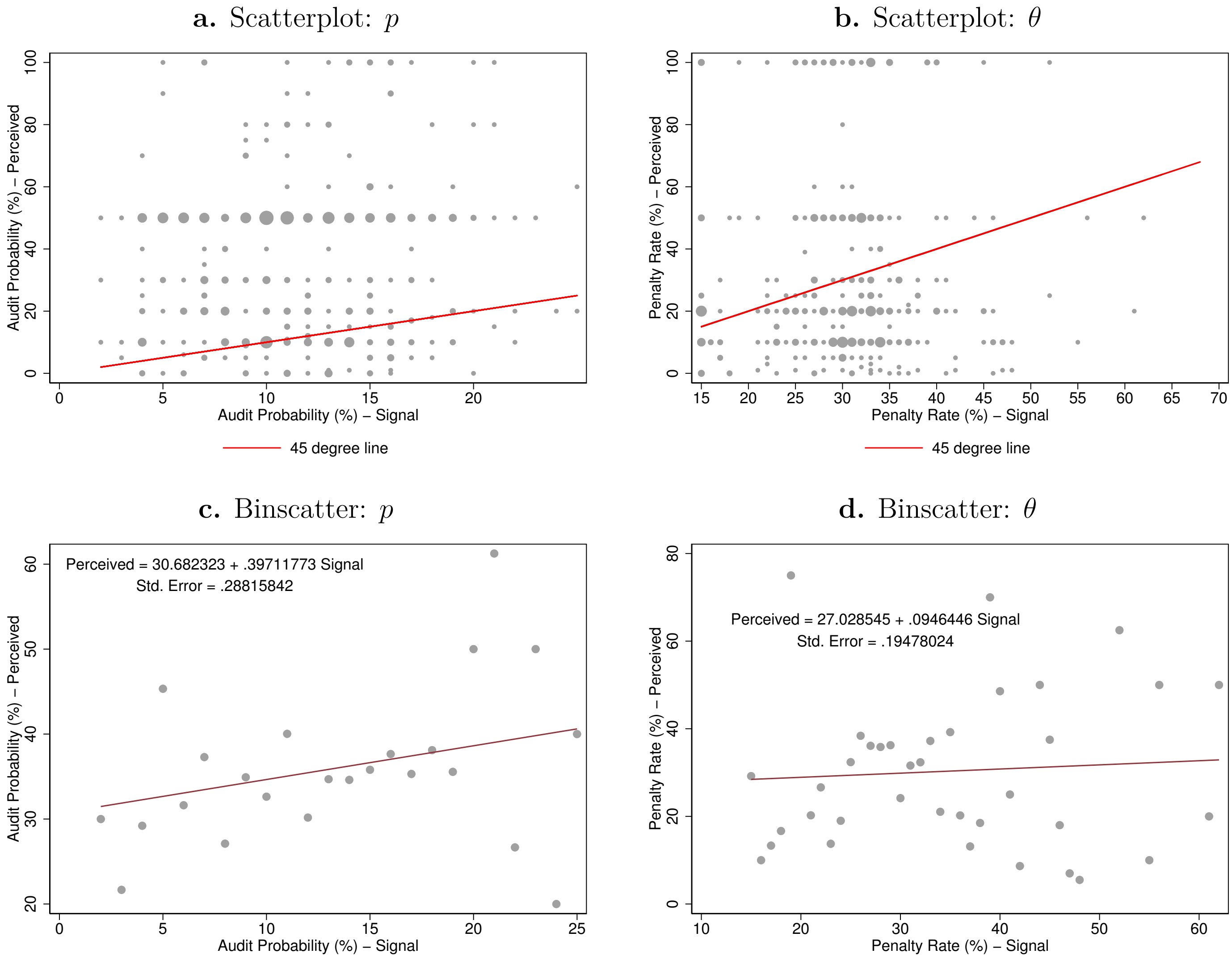
Notes: The histograms are based on the survey responses of those who, in the post-treatment survey, either self-identified as owners or did not answer the question regarding their occupation in the firm and those who did not answer 50% to the audit probability and penalty rate questions. "Perceived (pooled control group)" (N = 80) refers to survey respondents who received either the *baseline* (N = 46) or the *public-goods* (N = 34) letter during the experimental stage (neither of the two letters contained any information on audit probabilities or penalty rates). "Perceived (*audit–statistics*)" refers to respondents who received *audit-statistics* letters (N = 242). In panel (a), the x-axis represents the probability of being audited; in panel (b), it represents the average penalty rate. We report the mean responses and the p-value of the difference between the two groups. The answers correspond to Q2 and Q4 of the survey (see full survey questionnaire in Appendix A.7). The red line represents the density function of the information displayed in the *audit-statistics* letters, measured in the y-axis on the right (hidden for the sake clarity).

Figure B.6: Perception of Endogeneity of Audits: *Audit-Endogeneity* vs. *Baseline* and *Public-Goods*



Notes: "Perceived (pooled control group)" (N = 137) refers to respondents who received either the *baseline* or the *public-goods* letter during the experimental stage, while "Perceived Endogeneity" (N = 79) refers to respondents who received *audit-endogeneity* letters. These answers correspond to Q6 of the survey (see the full questionnaire in Appendix A.7). The x-axis represents the different categories presented as survey options. We report the mean responses for the two groups, measured on a five-point scale where one corresponds to "Increase a lot" and five to "Reduce a lot." We also report the p-value of the difference between the two groups.

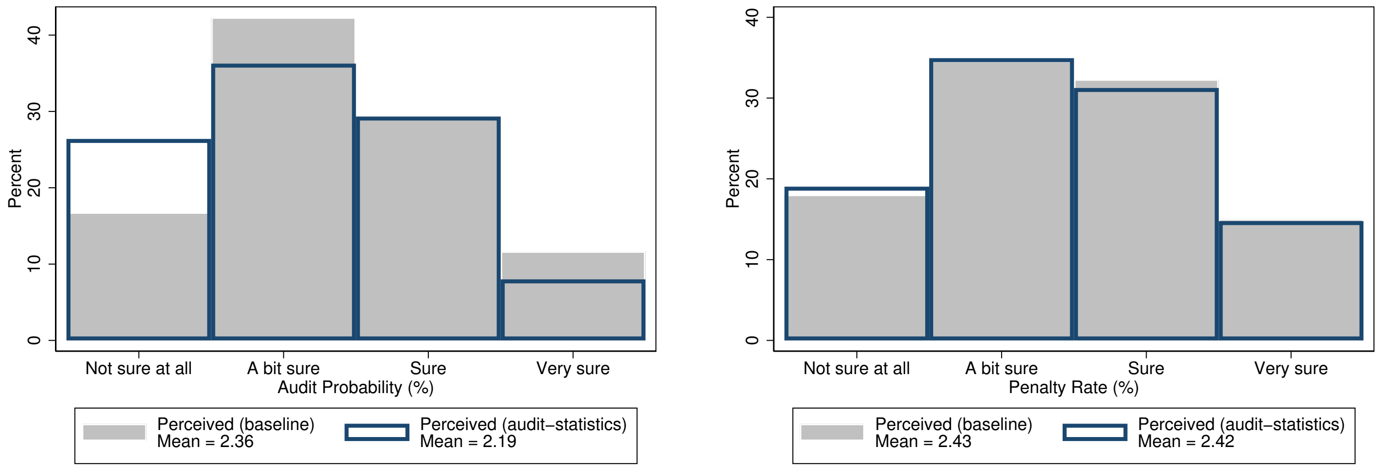
Figure B.7: Relation Between Signal Received and Self-Reported Perception



Notes: Figure B.7 shows the raw relations between the signal included in the letter and perceptions of audit probability and penalty rate reported in the survey for firms that received the *audit-statistics* letter (N = 365). The x-axis depicts the signal included in the letter and the y-axis the self-reported parameter. Panels (a) and (b) report the raw scatterplot where each dot represents a pair of values (signal, perception) and the size of the dot is proportional to the number of individuals. A solid 45-degree line is included in each panel. Panels (c) and (d) report the binned scatterplots of the same variable and include a line corresponding to the adjusted regression of perceptions over signals. Both the full equation predicted by the regression and the standard error of the coefficient on the signal are included as well. In all cases, the analysis is restricted to letters that were not returned by the postal service.

Figure B.8: Relation Between Signal Received and Self-Reported Certainty About Perception

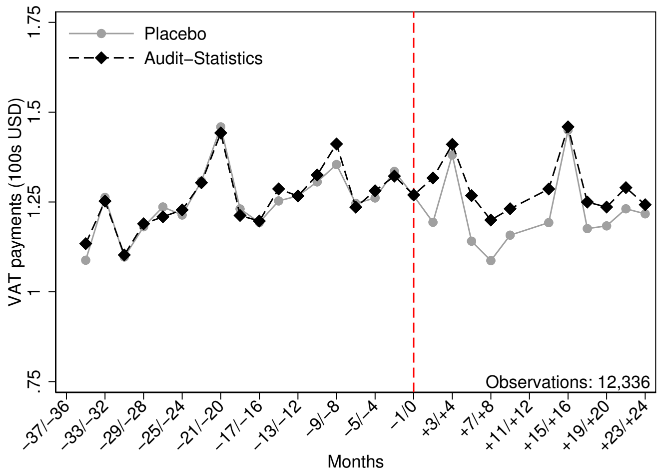
a. Certainty about Self-Reported p **b.** Certainty about Self-Reported θ



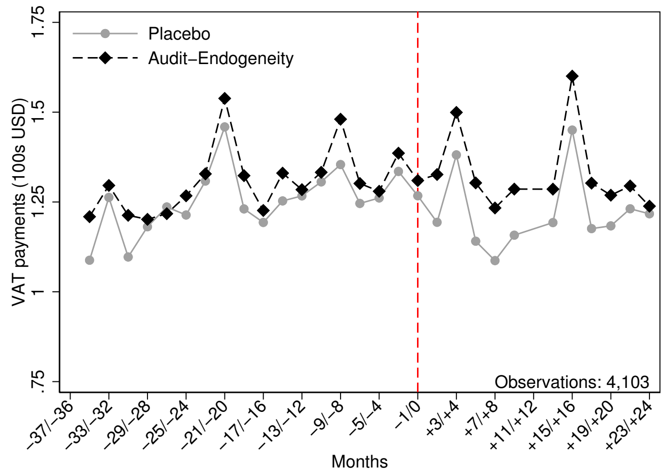
The histograms are based on the survey responses to Q3 and Q5 of those who, in the post-treatment survey, either self-identified as owners or did not answer the question regarding their occupation in the firm (see full survey questionnaire in Appendix A.7). "Perceived (*baseline)*" (N = 137) refers to the survey respondents who received the *baseline* letter (N = 68) or the *public-goods* letter (N = 69) during the experimental stage (neither of the two letters contained any information on audit probabilities or penalty rates). "Perceived (*audit–statistics*)" refers to respondents who received *audit-statistics* letters (N = 364). In both panels, the x-axis represents a four-point scale that captures individuals' degree of certainty with respect to perceived audit probability (panel (a)) and perceived penalty rate (panel (b)). We also report the mean responses and the p-value of the difference between the two groups, calculated on a scale from one (corresponding to the "Not at All Sure" response) to four (corresponding to the "Very Sure" response).

Figure B.9: Bimonthly VAT Payments, by Letter Type

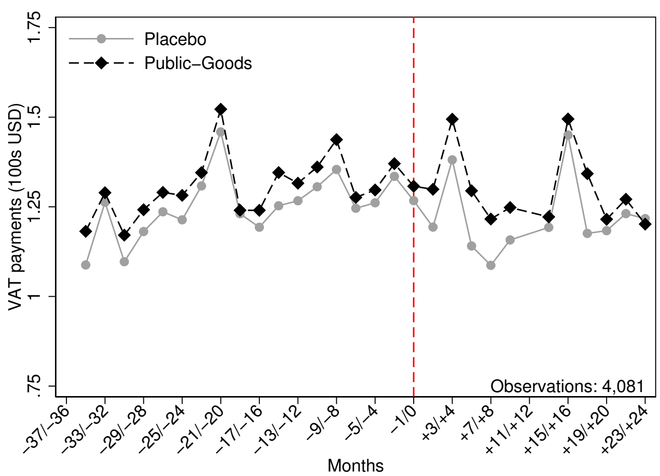
a. *Audit-Statistics*



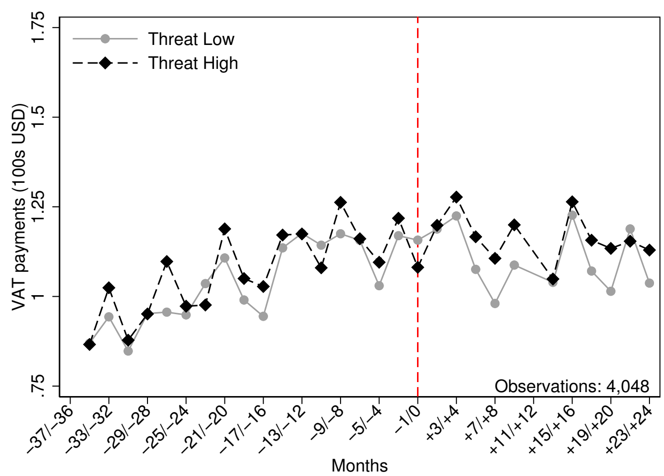
b. *Audit-Endogeneity*



c. *Public-Goods*



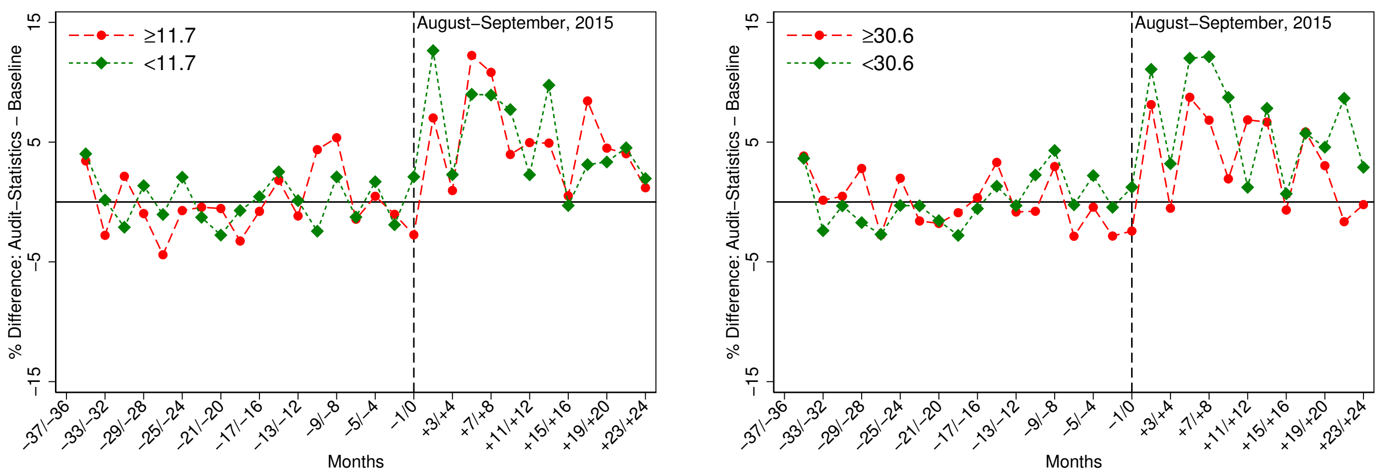
d. *Audit-Threat*



Notes: These figures plot the bimonthly total VAT payments separated by treatment and comparison groups. Similar figures for the difference between these two groups are reported in Figure 4. The data used for this figure correspond to the period from October 2012 to September 2017. The months of August and September 2015, when most of the letters were delivered, are defined as the reference bimonthly period and marked with the dashed vertical line. Each figure plots the time series of total VAT payments for the treatment arm and the comparison group separately. Panel (a) (N = 12,336) presents the evolution of the main outcome variable for firms that received the *audit-statistics* message and the *baseline* message. Panel (b) (N = 4,103) provides the results for the *audit-endogeneity* message, and panel (c) (N = 4,081) for the *public-goods* message. Panel (d) (N = 4,048) presents the two treatment groups selected from the secondary sample. In this case, the two lines depict the two *audit-threat* messages that contain different probabilities of being audited: 50% vs 25%. In all cases, analysis is restricted to letters that were not returned by the postal service. For each pair of months, VAT payments are top-coded at the 99.99th percentile to avoid contamination of the results by outliers. In all panels, we omit the +11/+12 pair of months (August 2016–September 2016) because the information provided by the IRS for it is incomplete.

Figure B.10: Effects of *Audit-Statistics* by Level of the Signal

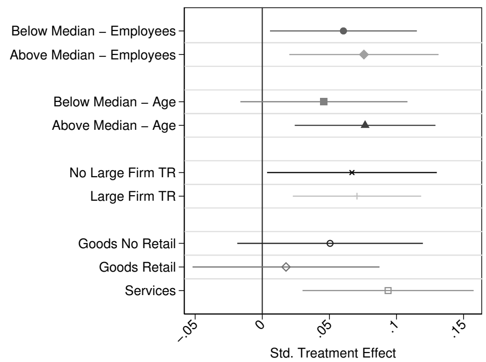
a. Effect of *audit-statistics* vs. *baseline*, by level of p **b.** Effect of *audit-statistics* vs. *baseline*, by level of θ



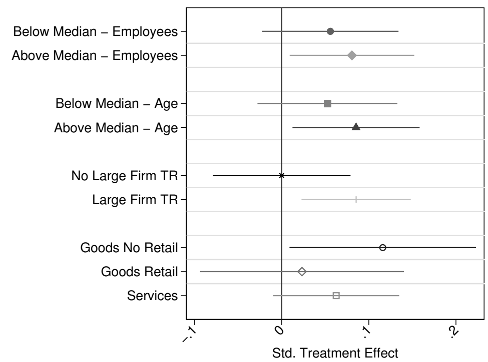
Notes: These figures plot the percentage difference in bimonthly total VAT payments between treatment and control groups, normalized by the average pre-treatment percentage difference (i.e., between months -35 and 0) for the same outcome, dividing the sample into firms that fall above and below the mean value of the corresponding signal. The red dots in panels (a) and (b) represent the effect for *audit-statistics* letter recipients with signals on p or θ above the mean respectively. Green dots represent the same effect but for those with signals on p and θ below the mean respectively. The data used in the figure cover the period October 2012–September 2017. The period of August 2015–September 2015, when most of the letters were delivered, is defined as the reference pair of months (it is indicated by the dashed vertical line). Each figure plots the difference between each treatment arm and the *baseline* letter (N = 12,336). In all cases, the analysis is restricted to letters that were not returned by the postal service. For each pair of months, VAT payments are top-coded at the 99.99th percentile to avoid the contamination of the results by outliers.

Figure B.11: Average Effects of *Audit-Statistics*, *Audit-Endogeneity,* and *Public-Goods* Letters on VAT Payments by Type of Firm

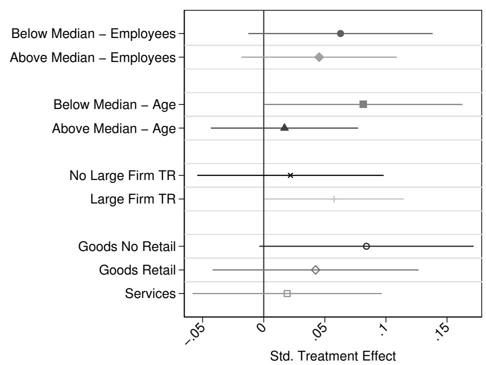
a. *Audit-Statistics*



b. *Audit-Endogeneity*



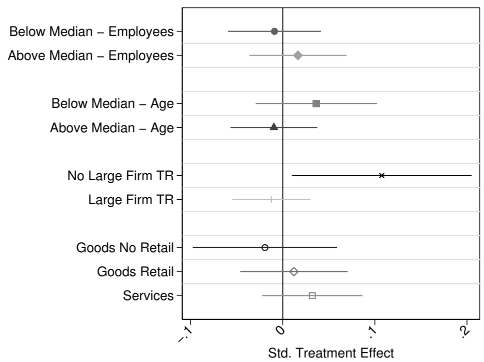
c. *Public-Goods*



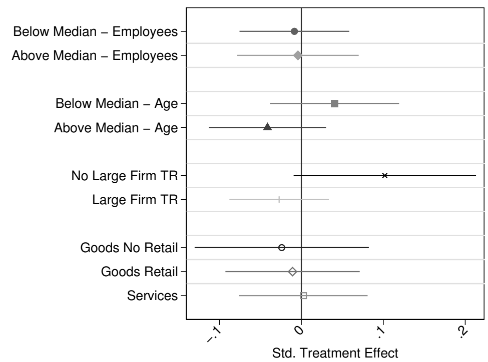
Notes: * significant at the 10% level, ** at the 5% level, *** at the 1% level. Standard errors are clustered at the firm level. Treatment effects are estimated for the first year post-treatment (October 2015–September 2016 vs. October 2014–September 2015) and based on an augmented version of the difference-in-differences specification reported in equation (1), which compares treated firms to control firms and pre-treatment to post-treatment periods using yearly aggregated variables. To capture the differential effects by group, we include a dummy variable for the interaction between the group of interest and the coefficient that captures the treatment effect. We also include all additional corresponding interactions with the time and treatment variables. More details about the baseline specification can be found in Table 2 and Section 4. More details about the augmented model can be found in Section B.8. Panel (a) compares the *audit-statistics* message with the *baseline* letter, while panels (b) and (c) replicate the comparison for the *audit-endogeneity* and *public-goods* messages respectively. The heterogeneity analysis is based on four characteristics: number of employees, age of firm (i.e., number of years registered with the IRS), type of tax return filed, and sector of operation. The median number of employees in the experimental sample is 2.5, while the median firm age is 12.7 years. Sixty-eight percent of firms filed a comprehensive VAT tax return in 2013 while the remaining 32% filed a simplified VAT tax return. Regarding the distribution of sector of operation, 29% of firms are non-retail sellers of goods , 22% are retail sellers, and 49% are service providers . In all cases, analysis is restricted to letters that were not returned by the postal service.

Figure B.12: Average Effects of *Audit-Statistics*, *Audit-Endogeneity,* and *Public-Goods* Letters on VAT Payments by Type of Firm, Pre-Treatment Falsification Test

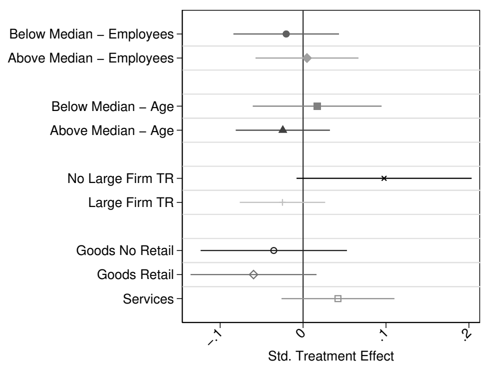
a. *Audit-Statistics*



b. *Audit-Endogeneity*



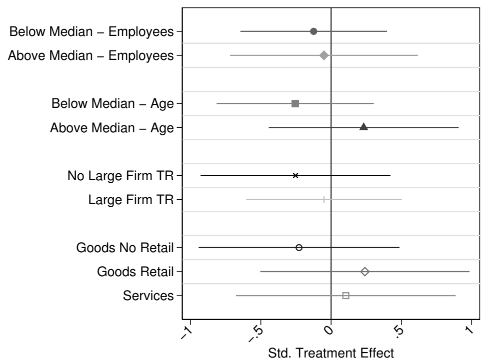
c. *Public-Goods*



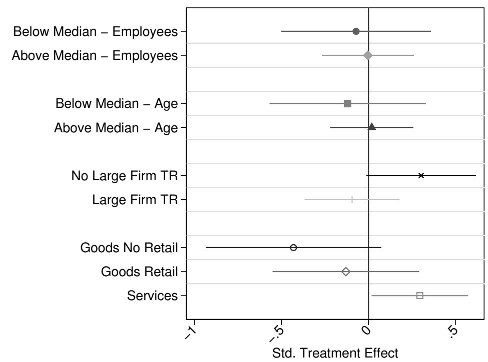
Notes: * significant at the 10% level, ** at the 5% level, *** at the 1% level. Standard errors are clustered at the firm level. Estimates presented in the figure correspond to falsification tests that compare two pre-treatment periods (October 2014–September 2015 vs. October 2013–September 2014) and are based on an augmented version of the difference-in-differences specification reported in equation (1), which compares treated firms to control firms using yearly aggregated variables. To capture the differential effects by group, we include a dummy variable for the interaction between the group of interest and the coefficient that captures the treatment effect. We include as well all additional corresponding interactions with the time and treatment variables. More details about the baseline specification can be found in Table 2 and Section 4. More details about the augmented model can be found in Section B.8. Panel (a) compares the *audit-statistics* message and the *baseline* letter, while panels (b) and (c) replicate the comparison for the *audit-endogeneity* and *public-goods* messages respectively. The heterogeneity analysis is based on four characteristics: number of employees, age of firm (i.e., number of years registered with the IRS), type of tax return filed, and sector of operation. The median number of employees in the experimental sample is 2.5, while the median age of the firms is 12.7 years. Sixty-eight percent of firms filed a comprehensive VAT tax return in 2013 while the remaining 32% filed a simplified VAT tax return. Regarding the distribution of sector of operation, 29% of firms are non-retail sellers of goods, 22% are retail sellers, and 49% are service providers. In all cases, analysis is restricted to letters that were not returned by the postal service.

Figure B.13: Elasticities of Tax Payments with Respect to Audit Probability and Penalty Rate in *Audit-Statistics* and *Audit-Threat* Sub-Treatments by Type of Firm

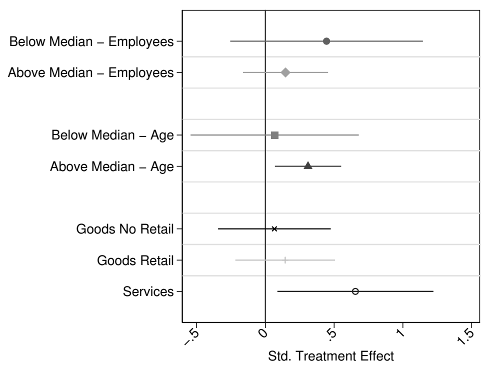
a. *Audit Probability*



b. *Penalty Rate*



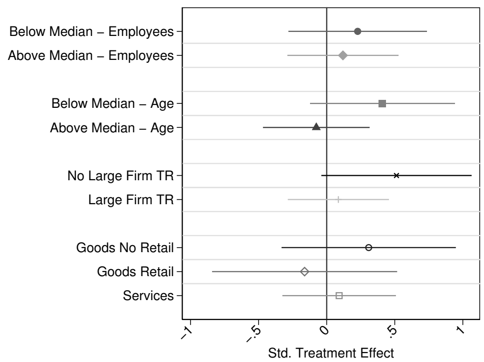
c. *Audit-Threat*



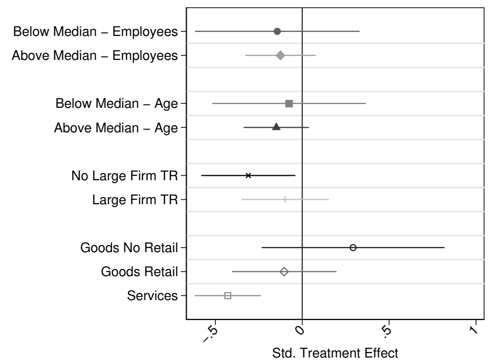
Notes: * significant at the 10% level, ** at the 5% level, *** at the 1% level. Standard errors are clustered at the firm level. Treatment effects are estimated for the first year post-treatment (October 2015–September 2016 vs. October 2014–September 2015) and are based on an augmented version of the difference-in-differences specification reported in equation (2) which compares treated firms that received different signals on p and θ. In all cases, we include an additional set of dummies for quintiles of the pre-treatment VAT payments, which are the groups from which we drew the sample of "similar firms" to calculate p and θ, and the corresponding interactions with the time variable. The results are based on Poisson regressions with variables expressed in percentage terms, so the coefficients can be interpreted directly as elasticities. To capture the differential effects by group, we include a dummy variable for the interaction between the group of interest and the coefficient that captures the treatment effect. We also include all additional corresponding interactions with the quintile, time, and treatment variables. More details about the baseline specification can be found in Table 3 and Section 5.2.1. More details about the augmented model can be found in Section B.8. Panel (a) presents the effect of providing different information regarding p in the *audit-statistics* message and panel (b) does the same for θ. Panel (c) compares the two *audit-threat* messages, i.e., the 50% threat of audit vs. the 25% threat of audit. Estimates should be interpreted as the effect of an additional percentage point of p and θ respectively in the information included in the letters on post-treatment VAT payments. The heterogeneity analysis is based on four characteristics: number of employees, age of the firm (i.e., the number of years registered with the IRS), type of tax return filed by the company, and sector of operation. The median number of employees in the experimental sample is 2.5, while the median age of the firms is 12.7 years. Sixty-eight percent of firms filed a comprehensive VAT tax return in 2013 while the remaining 32% filed a simplified VAT tax return. Regarding the distribution of sector of operation, 29% of firms are non-retail sellers of goods, 22% are retail sellers, and 49% are service providers. In all cases, analysis is restricted to letters that were not returned by the postal service.

Figure B.14: Elasticities of Tax Payments with Respect to Audit Probability and Penalty Rate in *Audit-Statistics* and *Audit-Threat* Sub-Treatments by Type of Firm, Pre-Treatment Falsification Test

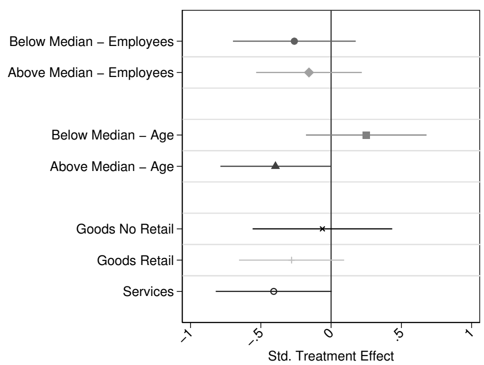
a. *Audit Probability*



b. *Penalty Rate*



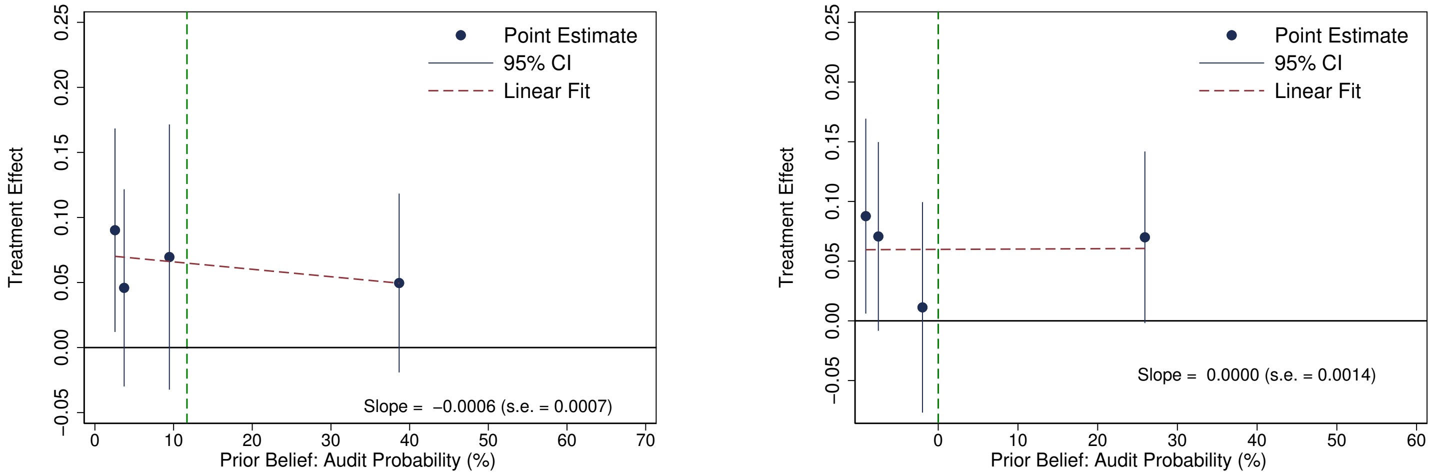
c. *Audit-Threat*



Notes: * significant at the 10% level, ** at the 5% level, *** at the 1% level. Standard errors are clustered at the firm level. Estimates presented in the figure correspond to falsification tests that compare two pre-treatment periods (October 2014–September 2015 vs. October 2013–September 2014) and are based on an augmented version of the difference-in-differences specification reported in equation (2), which compares treated firms that received different signals on p and θ. In all cases, we include an additional set of dummies for quintiles of the pre-treatment VAT payments, which are the groups from which we drew the sample of "similar firms" to calculate p and θ and the corresponding interactions with the time variable. The results are based on Poisson regressions with variables expressed in percentage terms, so the coefficients can be interpreted directly as elasticities. To capture the differential effects by group, we include a dummy variable for the interaction between the group of interest and the coefficient that captures the treatment effect. We include as well all additional corresponding interactions with the quintile, time, and treatment variables. More details about the baseline specification can be found in Table 3 and Section 5.2.1. More details about the augmented model can be found in Section B.8. Panel (a) presents the effect of providing different information regarding p in the *audit-statistics* message and panel (b) regarding θ. Panel (c) compares the two *audit-threat* messages, i.e., the 50% threat of audit vs. the 25% threat of audit. Estimates should be interpreted as the effect on post-treatment VAT payments of an additional percentage point of p and θ respectively in the information in the letters. The heterogeneity analysis is based on four characteristics: number of employees, age of firm (i.e., number of years registered with the IRS), type of tax return filed by the company, and sector of operation. The median number of employees in the experimental sample is 2.5, while the median age of the firms is 12.7 years. Sixty-eight percent of firms filed a comprehensive VAT tax return in 2013 while the remaining 32% filed a simplified VAT tax return. Regarding the distribution of sector of operation, 29% of firms are non-retail sellers of goods, 22% are retail sellers, and 49% are service providers. In all cases, analysis is restricted to letters that were not returned by the postal service.

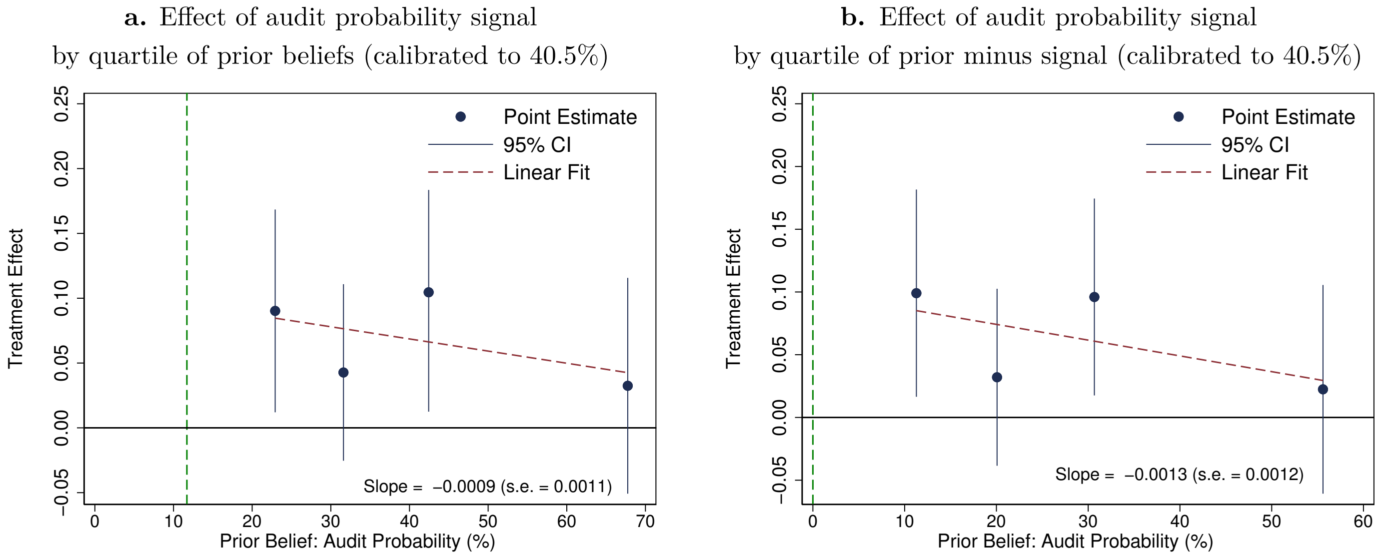
Figure B.15: Effect of *Audit-Statistics* vs. *Baseline* by Prior Beliefs

a. Audit probability - *prior (calibrated to 11.7%)* **b.** Audit probability - *prior - signal (calibrated to 11.7%)*



Notes: Panel (a) plots the first-year effect (October 2015–September 2016 vs October 2014–September 2015) of the *audit-statistics* letter on total VAT payments by quartiles of prior beliefs, while panel (b) reports the same results by quartiles of the difference between the prior belief and the signal sent in the *audit-statistics* message (N = 11,989). The prior belief, that is, belief before the experiment, is computed as $\hat{p}_i = 1 - \left(1 - \frac{\alpha_0 + N_i}{\alpha_0 + \beta_0 + T_i}\right)^3$ with parameters $\alpha_0 = 0.13$ and $\beta_0 = 1$ such that the mean prior belief about the probability of being audited at least once in the following three years matches the actual average probability observed in our sample. In panel (b), the signal for the placebo group was randomly assigned using the same strategy as for the *audit-statistics* group. The red dashed line represents the linear fit corresponding to the four estimates. In panel (a), the dashed vertical line represents the average actual probability of being audited as provided in the *audit-statistics* letters (11.7%). In panel (b), it represents the point at which the prior belief and the signal provided are equal. In both panels, each dot represents the estimated treatment effect for each quartile of the variable considered. Regressions are estimated using the baseline difference-in-differences specification reported in equation (1), but additional interaction terms are included for each quartile. More details are reported in the notes to Table 2 and Section 4. All effects are depicted with 95% confidence intervals. The results are based on Poisson regressions, which means the coefficients can be interpreted directly as semi-elasticities. Confidence intervals are computed with standard errors clustered at the firm level.

Figure B.16: Effect of *Audit-Statistics* vs. *baseline* by Prior Beliefs



Notes: Panel (a) plots the first-year effect (October 2015–September 2016 vs October 2014–September 2015) of the *audit-statistics* letter on total VAT payments by quartiles of prior beliefs, while panel (b) reports the same results by quartiles of the difference between the prior belief and the signal sent in the *audit-statistics* message (N=11,989). The prior belief is computed as $\hat{p}_i = 1 - \left(1 - \frac{\alpha_0 + N_i}{\alpha_0 + \beta_0 + T_i}\right)^3$ where $\alpha_0 = 1.36$ and $\beta_0 = 1$ such that the mean prior belief about the probability of being audited at least once in the following three years matches the average probability perceived by firms in the *baseline* and *public-goods* groups according to survey answers (40.5%). In panel (b), the signal for the placebo group was randomly assigned using the same strategy as for the *audit-statistics* group. The red dashed line represents the linear fit corresponding to the estimates for the four quartiles. In panel (a), the green dashed line represents the average perceived probability (40.5%) for the comparison group, and in panel (b) it represents the point at which the prior belief and the signal provided are equal. In both panels, each dot represents the estimated treatment effect for each quartile of the variable considered. Regressions are estimated using the baseline difference-in-differences specification reported in equation (1), but additional interactions terms are included for each quartile. More details are reported in the notes to Table 2 and Section 4. All effects are depicted with 95% confidence intervals. The results are based on Poisson regressions, which means the coefficients can be interpreted directly as semi-elasticities. Confidence intervals are computed with standard errors clustered at the firm level.

Table B.1: Comparison of Firm Characteristics for Different Groups

		Experimental Sample			
	All firms (1)	Main (2)	Secondary (3)	Invited to the survey (4)	p-value test Survey. vs Non-Survey (5)
Share paid VAT (3 months pre-mailing)	0.778	0.927	0.894	0.926	0.875
	(0.001)	(0.002)	(0.005)	(0.004)	
Amount of VAT paid (3 months pre-mailing)	3.717	1.894	1.744	1.894	0.993
	(0.033)	(0.022)	(0.067)	(0.048)	
Years registered in tax agency	14.208	15.233	19.437	14.445	0.001
	(0.039)	(0.132)	(0.202)	(0.159)	
Share audited between 2013-2015	0.065	0.102	0.141	0.079	<0.001
	(0.001)	(0.003)	(0.007)	(0.006)	
Number of employees	12.653	4.838	5.600	6.431	0.000
	(0.657)	(0.208)	(0.093)	(0.867)	
Share file comprehensive tax return in 2013	0.447	0.684	0.999	0.569	<0.001
	(0.001)	(0.004)	(0.000)	(0.008)	
Share no retail goods sector	0.246	0.288	0.432	0.263	<0.001
	(0.001)	(0.003)	(0.008)	(0.007)	
Share retail goods sector	0.132	0.216	0.328	0.150	<0.001
	(0.001)	(0.003)	(0.007)	(0.006)	
Share services sector	0.622	0.496	0.240	0.588	<0.001
	(0.001)	(0.004)	(0.007)	(0.008)	
N	120,142	16,392	4,048	3,845	

Notes: This table presents the average characteristics for different subsamples of the universe of firms registered with the tax agency (standard deviations in parentheses). Column (1) includes all firms that submitted at least one VAT payment in 2014 or 2015. Column (2) includes the subset of firms selected for the experimental sample according to the criteria described in section 3.2. Column (3) represents a group of high-risk firms selected by the IRS and that received the *audit-threat* letter. Column (4) corresponds to firms in the main sample with valid email addresses on file with the IRS. These are the firms that were selected to participate in the online survey conducted after the experiment. Column (5) reports the p-value of the means test between firms in the experimental sample that were invited to the survey and firms that were not. All data are based on administrative tax records (monthly payments, annual tax returns, and auditing registers), and Columns (2) to (4) are restricted to firms with successful letter delivery, which is the group of firms used in the empirical analysis. Robust standard errors are provided in parentheses.

Table B.2: Delivery Status by Treatment Arm

	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)
Delivered (%)	72.215	72.094	73.004	72.869	0.781	77.360	78.086	0.555
	(0.396)	(0.888)	(0.876)	(0.879)		(0.875)	(0.861)	
In Process (%)	0.219	0.117	0.234	0.274	0.660	0.131	0.303	0.211
	(0.041)	(0.068)	(0.095)	(0.103)		(0.076)	(0.114)	
Returned (%)	19.694	21.057	20.569	19.312	0.290	11.932	11.953	0.982
	(0.352)	(0.807)	(0.798)	(0.781)		(0.678)	(0.675)	
Missing (%)	7.873	6.732	6.194	7.545	0.011	10.577	9.658	0.302
	(0.238)	(0.496)	(0.476)	(0.522)		(0.643)	(0.615)	
N	12,791	2,555	2,567	2,558		2,288	2,309	

Notes: This table reports the delivery status of the letters by treatment arm. Columns (1) through (4) report the distribution of delivery statuses by treatment arm for firms in the main sample, while column (5) reports the p-value of the joint equality test. Columns (6) through (8) provide an analogous breakdown for the secondary sample. Information on delivery status is provided by the Uruguayan Post Office using the certified delivery service. "Delivered" letters are letters that were delivered by the post office and signed for by the recipient. "In Process" letters are letters that were in the process of being delivered, but not yet delivered, when the post office provided us with the delivery report. "Returned" letters are letters returned to the sender as the post office was not able to deliver them. Finally, there are some letters for which the post office did not provide any information. These are coded as "Missing."

Table B.3: 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	8.21	0.95	1.97	4.83	10.93	19.49
Retroactive VAT Amounts							
Post-treatment	0.30	1.40	0.00	0.00	0.00	0.00	0.62
Pre-Treatment	0.40	1.85	0.00	0.00	0.00	0.00	0.77
Concurrent VAT Amounts							
Post-treatment	6.16	7.51	0.33	1.16	3.52	8.07	15.84
Pre-Treatment	7.37	7.85	0.86	1.83	4.47	10.28	18.72
Other Taxes Amounts							
Post-treatment	3.30	5.43	0.00	0.95	1.81	3.52	7.42
Pre-Treatment	4.05	8.54	0.04	1.44	2.13	4.36	8.58
Total Taxes Amounts							
Post-treatment	9.77	11.31	1.04	2.68	6.14	12.39	23.33
Pre-Treatment	11.82	13.61	1.84	3.64	7.45	15.83	27.20
Some Payment							
Post-treatment	0.96	0.20	1.00	1.00	1.00	1.00	1.00
Pre-Treatment	1.00	0.06	1.00	1.00	1.00	1.00	1.00
More Than 6 payments							
Post-treatment	0.71	0.45	0.00	0.00	1.00	1.00	1.00
Pre-Treatment	0.89	0.32	0.00	1.00	1.00	1.00	1.00
More Than 3 payments							
Post-treatment	0.91	0.29	1.00	1.00	1.00	1.00	1.00
Pre-Treatment	0.98	0.15	1.00	1.00	1.00	1.00	1.00
VAT in Tax Return							
Post-treatment	7.19	9.93	0.00	0.49	3.84	9.69	18.97
Pre-Treatment	7.79	7.91	0.14	1.88	5.20	11.57	18.91

Notes: The statistics in this table correspond to firms that received the *baseline* letter (N = 2,064). The pre-treatment period spans from October 1, 2014 to September 30, 2015, and the post-treatment period from October 1, 2015 to September 30, 2016. "VAT Amounts" correspond to total VAT payments and withholdings (the sum of retroactive and concurrent VAT amounts). "Retroactive VAT Amounts" correspond to VAT payments and withholdings submitted for liabilities incurred two or more months prior to payment (e.g., VAT payments made in March 2016 for September 2015). "Concurrent VAT Amounts" include VAT payments made for the current or previous month. "Other Taxes Amounts" include payments of corporate income tax, wealth tax, and other taxes specific to a given business activity. "Total Taxes Amounts" is the sum of "VAT Amounts" and "Other Taxes Amounts." "VAT in Tax Return" is the final tax liability calculated in the yearly tax return that firms submit to the Uruguayan IRS.

Table B.4: Response Rates by Treatment Arm

	Audit Statistics (1)	Public Goods (2)	Audit Endogeneity (3)	Baseline (4)	Audit Threat (5)
a. Total firms by treatment arm					
Total	10,272	2,017	2,039	2,064	4,048
b. Invited to the survey					
Total	2,408	490	480	467	22
%	23.44	24.29	23.54	22.63	0.54
c. Started the survey					
Total	584	117	135	107	5
%	24.25	23.88	28.13	22.91	22.73
d. Owners that responded the survey					
Total	449	87	105	82	2
%	76.88	74.36	77.78	76.64	40.00
e. Non-Skip Rate - audit probability					
Total	351	64	82	65	1
%	92.61	92.75	95.35	95.59	100.00
f. Non-Skip Rate - penalty size					
Total	320	62	76	64	1
%	88.89	95.38	91.57	98.46	100.00

Notes: This table describes the composition of the online survey participants by treatment arm. Panel (a) reports the total number of firms by treatment arm. Panel (b) reports the raw number of invitations sent to firms in the experimental sample. The denominator for the percentage reported in this panel is the number of firms in each treatment arm reported in panel (a). Panel (c) reports the number of firms that started the survey (i.e., answered at least the first two questions). The denominator for the percentage reported in this panel is calculated using the totals reported in panel (b)as. Panel (d) reports the number of individuals who identified themselves as owners or who did not answer the question regarding their occupation in the firm, conditional on having started the survey (i.e., using the totals reported in panel (c) as the denominator). Finally, panels (e) and (f) report the non-skip rates–i.e., the response rate conditional on having answered the previous question–by treatment arm for the two key questions in the survey on perceptions of audit probability and penalty rate.

Table B.5: Effect of Treatment Status on Probability of Answering the Relevant Questions

	Answer: Perceived p		Answer: Perceived θ	
	Unconditional (1)	Conditional (2)	Unconditional (3)	Unconditional (4)
Treated	0.019	0.009	-0.010	-0.052
	(0.033)	(0.030)	(0.033)	(0.040)
Observations	1,105	267	1,105	267

Notes: * significant at the 10% level, ** at the 5% level, *** at the 1% level. Robust standard errors reported in parentheses. This table reports the effect of receiving the *audit-statistics* letter on probability of answering the two key questions in our survey analysis. The treatment variable is defined as one if the firm received the *audit-statistics* letter and zero if the firm received the *baseline* or the *public-goods* letter (our pooled control group). Column (1) reports the effect of receiving the *audit-statistics* letter on the probability of answering the audit probability question, as column (3) does for the penalty rate question. Because the treatment could have induced differential drop-out rates, Columns (2) and (4) report the same results but conditioning the sample to individuals who answered every single question up to the audit probability and penalty rates questions. We use OLS regressions to estimate the treatment effects. In all cases, the analysis is restricted to individuals who are owners of the firm or who did not answer the question regarding their occupation at it.

Table B.6: Balance Test for Survey Responses Conditional on Having Answered Perceptions Questions

	Main Sample				
	Audit Statistics (1)	Public Goods (2)	Audit Endogeneity (3)	Baseline (4)	p-value test (5)
Age	27.752	24.441	25.892	29.765	0.012
	(0.561)	(1.260)	(1.052)	(1.269)	
Female (%)	48.864	48.485	50.617	52.308	0.955
	(2.668)	(6.199)	(5.590)	(6.243)	
Capital city (%)	66.860	68.254	62.821	56.923	0.419
	(2.542)	(5.912)	(5.508)	(6.190)	
Very small firms (%)	11.966	9.231	14.634	15.152	0.685
	(1.735)	(3.618)	(3.927)	(4.447)	
Professional services (%)	50.829	44.118	44.578	52.941	0.540
	(2.631)	(6.066)	(5.489)	(6.098)	
Years registered with tax agency	16.366	13.385	15.025	17.894	0.036
	(0.539)	(0.925)	(1.033)	(1.221)	
Operates in one location (%)	61.582	62.687	64.634	67.164	0.829
	(2.589)	(5.953)	(5.312)	(5.781)	
One employee (%)	66.376	45.098	57.407	54.762	0.027
	(3.129)	(7.037)	(6.792)	(7.773)	
Audited in the last 3 years (%)	6.849	13.235	12.048	7.246	0.190
	(1.324)	(4.140)	(3.595)	(3.144)	
N	365	68	83	69	

Notes: This table reports the averages for different self-reported characteristics in the survey data by treatment group. 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. "Age," "Female," and "Capital city" refer to individual characteristics of the respondent. "Very small firms" refers to firms that participate in a special VAT regime with a fixed VAT annual payment. "Professional services" include individuals that pay VAT because they provide professional services. Observations included in the analysis are restricted to individuals who are owners of the firm or who did not answer the question regarding their occupation at it and who answered the two questions relevant to our analysis, that is, those regarding perceived audit probability and penalty rate.

Table B.7: Average Effects of *Audit-Statistics*, *Audit-Endogeneity,* and *Public-Goods* Messages: Alternative Specifications and Data Sources

	Made VAT Payments			VAT Payments			VAT Tax Returns
	≥ 1	≥ 3	≥ 6				Balance
	Probit (1)	Probit (2)	Probit (3)	Poisson (4)	OLS (5)	Tobit (6)	Poisson (7)
a. *Audit-Statistics* (10,272 firms [6,088]) vs *Baseline* (2,064 firms [1,270])							
Post-Treatment	0.095	-0.019	-0.045	0.070***	0.459***	0.451***	0.056*
	(0.132)	(0.064)	(0.040)	(0.021)	(0.147)	(0.153)	(0.029)
Pre-Treatment				0.009	0.069	0.052	0.007
				(0.020)	(0.154)	(0.155)	(0.032)
b. *Audit-Endogeneity* (2,039 firms [1,233]) vs *Baseline* (2,064 firms [1,270])							
Post-Treatment	0.109	-0.021	-0.039	0.071***	0.444**	0.452**	0.059
	(0.161)	(0.084)	(0.052)	(0.028)	(0.199)	(0.206)	(0.041)
Pre-Treatment				-0.005	-0.034	-0.048	0.011
				(0.028)	(0.219)	(0.220)	(0.040)
c. *Public-Goods* (2,017 firms [1,240]) vs *Baseline* (2,064 firms [1,270])							
Post-Treatment	-0.099	-0.140	-0.116**	0.051**	0.301*	0.332*	-0.021
	(0.200)	(0.091)	(0.053)	(0.025)	(0.177)	(0.184)	(0.038)
Pre-Treatment				-0.003	-0.019	-0.015	0.015
				(0.024)	(0.186)	(0.187)	(0.039)

Notes: * significant at the 10% level, ** at the 5% level, *** at the 1% level. Standard errors reported in parentheses are clustered at the firm level. Treatment effects are estimated using the difference-in-differences specification reported in equation (1), which compares treated firms to control firms and pre-treatment to post-treatment periods using yearly aggregated variables. In the first row of each panel ("Post-Treatment"), the coefficient corresponds to a comparison between a post-treatment period and a pre-treatment period (October 2015–September 2016 vs October 2014–September 2015). The second row ("Pre-Treatment") presents a falsification test where two pre-treatment periods are compared (October 2014–September 2015 vs October 2013–September 2014). Additional details about the model are reported in the notes to Table 2 and Section 4. Panel (a) compares the *audit-statistics* message with the *baseline* letter, while panels (b) and (c) replicate the comparison for the *audit-endogeneity* and *public-goods* messages respectively. Column (1) presents the treatment effect on the probability of making at least one VAT payment in the post-treatment period using a Probit model. Columns (2) and (3) are identical to column (1), except that the dependent variable is defined as making at least three or six payments in the post-treatment period respectively. Columns (4), (5), and (6) present different estimation strategies for the intensive margin, i.e., the total amount of VAT paid. Column (4) corresponds to the baseline specification (a Poisson model), while column (5) uses an OLS regression and column (6) a Tobit regression. Column (7) reports the result of estimating a Poisson model on the final VAT liability calculated from annual tax returns. The number of observations corresponding to specifications reported in column (7) is reported between brackets.

Table B.8: Effects of *Audit-Statistics* and *Audit-Threat* Sub-Treatments: Alternative Specifications

	Made VAT Payments			VAT Payments			VAT Tax Returns
	≥ 1	≥ 3	≥ 6				Balance
	Probit (1)	Probit (2)	Probit (3)	Poisson (4)	OLS (5)	Tobit (6)	Poisson (7)
a. *Audit-Statistics* (N=10,272 firms [6,088])							
Audit Probability (%)							
Post-Treatment	0.205	0.481	0.121	-0.063	-0.517	-0.128	-0.218
	(1.333)	(0.669)	(0.380)	(0.242)	(1.783)	(1.851)	(0.257)
Pre-Treatment				0.141	1.141	1.190	-0.038
				(0.164)	(1.336)	(1.343)	(0.230)
Penalty Size (%)							
Post-Treatment	0.717	-0.353	-0.005	-0.033	-0.233	-0.152	-0.223
	(0.575)	(0.333)	(0.202)	(0.118)	(0.976)	(0.995)	(0.138)
Pre-Treatment				-0.128	-1.130	-1.148	-0.092
				(0.108)	(0.964)	(0.970)	(0.114)
b. *Audit-Threat* (N=4,048 firms [3,236])							
Post-Treatment	0.044	0.229	0.007	0.217	1.375	1.298	0.354**
	(0.435)	(0.240)	(0.170)	(0.142)	(0.901)	(0.997)	(0.178)
Pre-Treatment				-0.185	-1.215	-1.246	-0.249
				(0.157)	(1.072)	(1.117)	(0.167)

Notes: * significant at the 10% level, ** at the 5% level, *** at the 1% level. Standard errors reported in parentheses are clustered at the firm level. Treatment effects are estimated using the difference-in-differences specification reported in equation (2), which compares treated firms that received different signals on p and θ. In all cases, we include an additional set of dummies for quintiles of the pre-treatment VAT payments, which are the groups from which we drew the sample of "similar firms" to calculate p and θ and the corresponding interactions with the time variable. The results are based on Poisson regressions with variables expressed in percentage terms, which means the coefficients can be interpreted directly as elasticities. 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. In the "Post-Treatment" rows, the coefficient reported corresponds to a comparison between a post-treatment period and a pre-treatment period (October 2015–September 2016 vs October 2014–September 2015). In the "Pre-Treatment" rows, we present a falsification test where two pre-treatment periods are compared (October 2014–September 2015 vs October 2013–September 2014). Columns (1) and (2) show the 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 Probit model. Columns (2) and (3) are identical to column (1), except that the dependent variable is defined as making at least three or six payments in the post-treatment period respectively. Columns (4), (5), and (6) present different estimation strategies for the intensive margin, i.e., the total amount of VAT paid. Column (4) corresponds to the baseline specification (a Poisson model), while column (5) uses an OLS regression and column (6) uses a Tobit regression. Column (7) reports the result of estimating a Poisson model on the final VAT liability calculated from annual tax returns. The number of observations corresponding to specifications reported in column (7) is reported in square brackets.

Table B.9: 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)
$p \cdot \theta$ (%)						
Post-Treatment	-0.292	-0.516	2.503	-0.406	0.398	0.017
	(0.529)	(0.560)	(3.095)	(0.532)	(0.531)	(0.439)
Pre-Treatment	-0.011	-0.011	-3.193	0.132	-0.173	-0.094
	(0.415)	(0.415)	(2.701)	(0.412)	(0.528)	(0.355)

Notes: * significant at the 10% level, ** at the 5% level, *** at the 1% level. Standard errors reported in parentheses are clustered at the firm level. Treatment effects are estimated using the difference-in-differences specification reported in equation (B.1), which compares treated firms that received different signals on $p \times \theta$. In all cases, we include an additional set of dummies for quintiles of the pre-treatment VAT payments, which are the groups from which we drew the sample of "similar firms" to calculate p and θ and the corresponding interactions with the time variable. This table presents the effect of providing different information regarding $p \times \theta$ in the *audit-statistics* message. In the "Post-Treatment" row, the coefficient reported corresponds to a comparison of a post-treatment period and a pre-treatment period. In the "Pre-Treatment" row, we present a falsification test where two pre-treatment periods are compared. Columns (1) and (2) report the effect of treatment by time horizon. The post-treatment effect reported in column (1) corresponds to the difference-in-differences estimate that compares October 2015–September 2016 vs. October 2014–September 2015. The post-treatment effect reported in column (2) is analogous but uses the second year after the treatment as the post-treatment period (i.e., October 2016 – September 2017). For the falsification tests, column (1) is based on a comparison of October 2014 – September 2015 and October 2013 – September 2014 while column (2) compares October 2014–September 2015 and October 2012–September 2013. Columns (3) and (4) present the first-year effect of treatment on retroactive (3) and concurrent (4) VAT payments. Columns (5) and (6) report the first-year results by type of tax. Column (5) presents the effect of the treatment on other (non-VAT) tax payments, while column (6) reports its effect on the total amount of taxes paid by the firms during the same period. In all cases, we restrict analysis to firms that effectively received the letter as reported by the postal service.

Table B.10: Average Effects of *Audit-Statistics*, *Audit-Endogeneity,* and *Public-Goods* Messages: Decomposition of Taxes

	Total Taxes (1)	VAT (2)	Other (3)	Other Taxes: Corporate (4)	Other Taxes: Wealth (5)	Other Taxes: PIT Withholdings (6)
a. *Audit-Statitstics* (10,272 firms) vs *Baseline* (2,064 firms)						
Post-Treatment	0.073***	0.070***	0.086**	0.058*	0.198	0.078
	(0.020)	(0.021)	(0.037)	(0.030)	(0.129)	(0.048)
Pre-Treatment	0.014	0.009	0.008	0.024	-0.057	0.030
	(0.021)	(0.020)	(0.043)	(0.030)	(0.166)	(0.046)
b. *Audit-Endogeneity* (2,039 firms) vs *Baseline* (2,064 firms)						
Post-Treatment	0.078***	0.071***	0.090*	0.069	0.204	-0.003
	(0.028)	(0.028)	(0.054)	(0.054)	(0.133)	(0.077)
Pre-Treatment	0.017	-0.005	0.056	0.063	0.019	0.072
	(0.028)	(0.028)	(0.055)	(0.049)	(0.169)	(0.080)
c. *Public-Goods* (2,017 firms) vs *Baseline* (2,064 firms)						
Post-Treatment	0.056**	0.051**	0.067	0.058	0.135	0.005
	(0.024)	(0.025)	(0.043)	(0.043)	(0.129)	(0.059)
Pre-Treatment	-0.015	-0.003	-0.038	-0.045	-0.005	-0.008
	(0.026)	(0.024)	(0.054)	(0.053)	(0.162)	(0.065)

Notes: * significant at the 10% level, ** at the 5% level, *** at the 1% level. Standard errors reported in parentheses are clustered at the firm level. Treatment effects are estimated using the difference-in-differences specification reported in equation (1), which compares treated firms to control firms and pre-treatment to post-treatment periods, using yearly aggregated variables. The results are based on Poisson regressions, which means the coefficients can be interpreted directly as semi-elasticities. Panel (a) compares the *audit-statistics* message to the *baseline* message, while panels (b) and (c) replicate the comparison for the *audit-endogeneity* and *public-goods* messages respectively. In the first row of each panel ("Post-Treatment"), the coefficient reported corresponds to a comparison of a post-treatment period (October 2015–September 2016) and a pre-treatment period (October 2014–September 2015). The second row ("Pre-Treatment") presents a falsification test where two pre-treatment periods are compared (October 2014–September 2015 and October 2013–September 2014). Column (1) reports the treatment effects on total tax payments. Columns (2) and (3) break down total tax payments into VAT and other tax payments. Columns (4), (5), and (6) break down other taxes into corporate income tax, property taxes, and personal income tax withholding respectively. In all cases, we restrict the analysis to firms that effectively received the letter as reported by the postal service.

Table B.11: Elasticities of Tax Payments with Respect to Audit Probability and Penalty Rate, *Audit-Statistics* and *Audit-Threat* Sub-Treatments: Breakdown of Taxes

				Other Taxes		
	Total Taxes (1)	VAT (2)	Other (3)	Corporate (4)	Wealth (5)	PIT Withholdings (6)
a. *Audit-Statitstics* (10,272 firms)						
Audit Probability (%)						
Post-Treatment	0.038	-0.063	0.109	0.316	-0.125	-0.288
	(0.208)	(0.242)	(0.240)	(0.293)	(0.358)	(0.573)
Pre-Treatment	0.063	0.141	-0.035	0.002	-0.609	0.517
	(0.147)	(0.164)	(0.230)	(0.257)	(0.652)	(0.509)
Penalty Size (%)						
Post-Treatment	-0.001	-0.033	0.061	0.001	0.021	0.608**
	(0.092)	(0.118)	(0.103)	(0.123)	(0.184)	(0.299)
Pre-Treatment	-0.078	-0.128	0.018	-0.066	0.450	-0.129
	(0.087)	(0.108)	(0.119)	(0.132)	(0.283)	(0.281)
b. *Audit-Threat* (4,048 firms)						
Audit Probability (%)						
Post-Treatment	0.233**	0.217	0.002	-0.060	0.146	2.567***
	(0.111)	(0.142)	(0.176)	(0.210)	(0.146)	(0.989)
Pre-Treatment	-0.257	-0.185	-0.067	-0.047	-0.100	-1.480
	(0.164)	(0.157)	(0.148)	(0.181)	(0.194)	(1.013)

Notes: * significant at the 10% level, ** at the 5% level, *** at the 1% level. Standard errors reported in parentheses are clustered at the firm level. Treatment effects are estimated using the difference-in-differences specification reported in equation (2), which compares treated firms that received different signals on p and θ. In all cases, we include an additional set of dummies for quintiles of the pre-treatment VAT payments, which are the groups from which we drew the sample of “similar firms” to calculate p and θ and the corresponding interactions with the time variable. The results are based on Poisson regressions with variables expressed in percentage terms, which means the coefficients can be interpreted directly as elasticities. 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. In the first row of each panel (“Post-Treatment”), the coefficient reported corresponds to a comparison of a post-treatment period (October 2015–September 2016) and a pre-treatment period (October 2014–September 2015). The second row (“Pre-Treatment”) presents a falsification test where two pre-treatment periods are compared (October 2014–September 2015 and October 2013–September 2014). Column (1) reports the treatment effects on total tax payments. Columns (2) and (3) break down total tax payments into VAT and other tax payments. Columns (4), (5), and (6) break down other taxes into corporate income tax, property taxes, and personal income tax withholding respectively. In all cases, we restrict the analysis to firms that effectively received the letter as reported by the postal service.

Table B.12: Effects of *Audit-Statistics* by Prior Beliefs ($\hat{p}$) and Information Treatment's Audit Probability (p), and by Previous Audit Experience

		By $\hat{p}$		Audited in 2001-2015	
	All (1)	$\hat{p} < p$ (2)	$\hat{p} \geq p$ (3)	Yes (4)	No (5)
a. *Audit-Statitstics* (10,272 firms) vs *Baseline* (2,064 firms)					
Post-Treatment	0.070***	0.061***	0.086***	0.082**	0.066***
	(0.021)	(0.023)	(0.028)	(0.039)	(0.025)
Pre-Treatment	0.009	0.011	0.004	-0.028	0.023
	(0.020)	(0.020)	(0.024)	(0.043)	(0.022)

Notes: * significant at the 10% level, ** at the 5% level, *** at the 1% level. Standard errors reported in parentheses are clustered at the firm level. Treatment effects are estimated using the difference-in-differences specification reported in equation (1), which compares treated firms to control firms and pre-treatment to post-treatment periods, using yearly aggregated variables. In the first row ("Post-Treatment"), the coefficient reported corresponds to a comparison between a post-treatment period and a pre-treatment period (October 2015–September 2016 vs October 2014–September 2015). The second row ("Pre-Treatment") presents a falsification test where two pre-treatment periods are compared (October 2014–September 2015 vs October 2013–September 2014). Additional details about the model are reported in the notes to Table 2 and Section 4. Column (1) shows results from the main specification, i.e., the first year post-treatment effect estimates reported in Table 2. Column (2) reports the effect of the *audit-statistics* message on firms whose prior beliefs about probability of being audited were below the signal (p) reported in the letter they received. Column (3) reports the results for firms whose prior beliefs were above the reported p. In both cases, the reference group is the *baseline* group. Column (4) reports estimates for firms that were audited at least once between 2001 and 2015, and column (5) for the group of firms not audited during that period.

C Solution to the Model

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 is:

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

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

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:

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

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

The closed-form expressions are still available for the extensions with $\alpha<1$ and $\epsilon>0$, but not for $p_1>0$ (we thus use standard numerical methods to compute these elasticities).

D Additional Material

Figure D.1: Mass Advertising Campaign Sample: Billboard Poster, United Kingdom's Tax Authority, 2012

Notes: Advertising campaign by the United Kingdom's tax authority, HMRC (Her Majesty's Revenue and Customs), 2012. Previously hosted at http://www.gov.uk/sortmytax (no longer available, accessed through http://web.archive.org/)