Can transparency hurt? An experiment on whether disclosure of audit policy details reduces tax compliance

Tan, Fangfang and Yim, Andrew

Max Planck Institute for Tax Law and Public Finance, Munich, Germany, Cass Business School, City University London

27 May 2011

Can Transparency Hurt? An Experiment on Whether Disclosure of Audit Policy Details Reduces Tax Compliance

Fangfang Tan

Max Planck Institute for Tax Law

and Public Finance, Munich, Germany

Tel: +49-89-24246-5252 Fax: +49-89-24246-5299

Email: tff626@gmail.com

Andrew Yim

Cass Business School

City University London, London, UK

Tel: +44-20-7040-0933 Fax: +44-20-7040-8881

Email: a.yim@city.ac.uk / andrew.yim@aya.yale.edu

This version: May 27, 2011

Abstract

Tax authorities around the world often are reluctant to disclose audit policy details. In particular, the US Internal Revenue Service (IRS) has the practice of releasing broad statistics like the audit rate of each income class but resists pressures demanding details on how different circumstances might result in a higher audit probability to taxpayers. This paper experimentally examines whether disclosing such details can reduce tax compliance. We compare a *Flat-rate* treatment, where taxpayers are told about the average audit probability, with a *Bounded* treatment, where taxpayers are fully informed of the contingent audit probability structure. Our findings do not support the potential concern against disclosing details. In an additional *Bounded*-hi-q treatment where multiple equilibria exist, the compliance level is even higher under full disclosure of the probability structure.

JEL codes: H26, M42, C9, C72

Keywords: Information disclosure, Government transparency, Audit policy, Tax auditing, Tax compliance, Laboratory experiment

Acknowledgements

We are particularly grateful for the valuable suggestions by Ralph-C Bayer, Luc Bissonnette, Eddy Cardinaels, Vilen Lipatov, Wieland Mueller and Charles Noussair. We thank James Alm, Cedric Argenton, Rudolf Avenhaus, Jochen Bigus, Kim Bloomquist, Folkert Botma, Liesbeth Bruynseels, Douglas DeJong, Lata Gangadharan, Karla Johnstone, Steven Kachelmeier, William Kinney, Tobias Klein, Ben Li, Ksenia Panidi, Jan Potters, Ernesto Reuben, Jason Shachat, Joel Slemrod, Bin Srinidhi, Konrad Stahl, Jeroen Suijs, Stefan Trautmann, Peter van Oudheusden, Zongxin Qian, Arthur van Soest, Gijs van de Kuilen, Ann Vanstraelen, Eyal Winter, Shlomo Yitzhaki and seminar participants of Tilburg University, the GSS interdisciplinary workshop at CentER, 2010 Behavioral Decision Research in Management (BDRM) Conference held at Carnegie Mellon University and 66th Annual Congress of the International Institute of Public Finance (IIPF) for helpful comments. We thank Chung Kim and Michael McKee for sharing their experimental instructions, and CentER lab for financial support. Siqi Pan and Huojun Sun provided excellent research assistance. This paper was partly written while Fangfang Tan was studying at Tilburg University and New York University, whose hospitality and financial support is greatly acknowledged. All remaining errors are ours.

1 Introduction

This paper asks whether more transparency in government agencies compromises their commissioned objectives. Specifically, we study the impact of information disclosure, concerning audit policy details, by a tax authority like the US Internal Revenue Service (IRS) on the level of tax compliance. IRS has long been accused of having a "secret culture" (see Saxon (1994), Johnston (1995b), and Davis (1997)).[1] While the agency is not as opaque as before, what people know about IRS audits is still mainly from broad statistics provided on its web site (e.g., from the IRS Data Books).[2] Even though, following the enactment of the IRS Restructuring and Reform Act of 1998, the agency has made public the Internal Revenue Manual (IRM) describing the tax audit process (Gates (2000)), certain details of the audit policy remain undisclosed to taxpayers.[3]

Why does IRS disclose only broad statistics like the audit rate of each income class but not details of the audit policy? Apparently, the agency worries that the tax compliance level would fall should taxpayers know details of the audit policy (*New York Times* (1981a) and *New York Times* (1981b)). In this paper, we investigate whether a tax authority could

[1]For example, "the [US] Government's chief keeper of historic records said [on 20 December 1995] that the Internal Revenue Service has, for at least two decades, violated Federal laws that require it to identify significant documents and turn them over to the National Archives. ... John W. Carlin, the Archivist of the United States, gave the I.R.S. 90 days to come up with a plan to identify, safeguard and eventually turn over to his office records that may have historic value. His 50-page evaluation cited "serious shortcomings" in I.R.S. record-keeping and questioned whether some important records had been lost or destroyed. "Numerous records that document both policy-making and high-profile programs" either are not scheduled to be released to the National Archives "or have not been located and identified," the evaluation said. ... Critics have long accused the I.R.S. of excessive secrecy, and historians, individual taxpayers and others have battled for access to statistical data." (Johnston (1995b))

[2]Documentation of IRS audit practices in the academic literature is sparse. An example is Pentland and Carlile (1996).

[3]Evidence for this is not hard to find. For example, the actual operation of the discriminant function (DIF) formula used to identify the most suspicious tax returns for follow-up remains a "closely guarded secret" (Jones (2001)). It seems that IRS has released the entire IRM on its web site. But a careful look at the web manual shows that the section IRM 4.19.1.2.6, Form 1040 Individual Returns Scored by DIF System (Audit Code Definitions) is missing (see the Table of Contents of Part 4 of the IRM at http://www.irs.gov/irm/part4/index.html). According to IRM 4.1.3.2.2, the missing section contains Audit Codes used to identify returns "delivered [directly] to Examination as Automatics for manual screening," regardless of the DIF scores. Apparently, IRS wants to keep the information secret.

Thoroughly searching over the web IRM can locate partial information about the Codes, e.g., in Exhibit 3.11.3-5. However, like multiple places of the section IRM 3.11.3 Individual Income Tax Returns (e.g., Exhibit 3.11.3-8 Examples of Reasonable Causes and Exhibit 3.11.3-6 Unallowable Codes), some details have been overwritten with equal signs (=). Similar blacked-out's can be found in other IRS documents released to the public, e.g., pages 3-4, 3-8, 3-14, and 3-15 of the 2010 version of "IRS Processing Codes and Information" (IRS (2010)).

be no worse off by fully disclosing to taxpayers the structure of an audit policy, instead of merely telling them the average audit probability. Answering this question is important. It can provide evidence to support the information-withholding position of IRS, or otherwise give some assurance to the agency to become more transparent, as critics have demanded, without compromising the objective to increase tax compliance.

The tension between increasing government transparency and keeping appropriate levels of secrecy is not new (Ginsberg (2011)). Watchdog organizations like OMB Watch, Citizens for Responsibility and Ethics in Washington, First Amendment Coalition, and Taxpayers for Common Sense always press for more transparency and freedom of information. However, the demand for government transparency has never been stronger (Ornstein and Limor (2011)).

Since Barack Obama was elected as the US President, the administration has emphasized the commitment to "creating an unprecedented level of openness in Government" (Obama (2009)). As a result of the Open Government Initiative, agencies are asked to increase disclosures (see, e.g., Department of the Treasury (2011)). Still, censorship of information prior to release is not unheard of (The Associated Press (2011)).

As for IRS, the reluctance to disclose information has not changed much in the last four decades.[4] The reservation is not only on open disclosure to the public but also on confidential

[4]In 1973, Ralph Nader of Tax Reform Research Group invoked the Freedom of Information Act (FOIA) in order to obtain some IRS documents. The agency refused and Nader responded with a suit before IRS reluctantly agreed to provide the documents (*Time* (1974)). A year before this, Susan Long and her husband started the litigation lasting for over 15 years, involving courts as high as the US Supreme Court, forcing the agency to be more open in releasing information. "Their first successful legal action set the principle that the I.R.S. could not withhold information like statistics on the audit rates for taxpayers in different income groups, nor its basic operating manual." (Saxon (1994)).

It was thought that the nearly two decades of litigation finally came to an end when IRS was ordered to pay the Longs' legal fees in 1991. But the battle reopened in 2004 when IRS told Susan Long that after extensive research, its lawyers concluded that "no court order existed and 'accordingly, the I.R.S. is not in violation of any standing injunctions'" by withholding information from her (Johnston (2006)). In 2006, Long went to court again to file a legal motion to require the agency to comply with prior court orders to turn over detailed data on its audit practices (Johnston (2006)). On 13 June 2008, the US District Court in Seattle granted her motion. IRS timely appealed the order. Finally, on 16 September 2010, the US Court of Appeals for the Ninth Circuit affirmed in part and reversed in part the 13 June 2008 order, ruling that some information taken from IRS's Form 5344 of one particular taxpayer, referred to as "cells of one," is confidential under 26 U.S. Code Section 6103(b) (United States Courts of Appeals for the Ninth Circuit (2010)).

Long was not the only one in battle with IRS for information disclosure in recent years. For example, Tax Analysts, the nonprofit publisher of *Tax Notes* magazine, went to court for obtaining e-mail messages in which tax auditors in the field were given advice on how to apply the law. "We won a unanimous court of appeals decision that they can't hide this stuff," Tax Analysts' president said, "but instead of complying with the order to produce it, they are playing games." (Johnston (2008)).

disclosure to researchers (see, e.g., Shackelford and Shevlin (2001), page 375). Intrigued by the puzzling attitude of IRS, we are interested in verifying whether disclosing audit policy details necessarily reduce tax compliance, or maybe it could actually increase compliance.

Laboratory tightly controls many factors that may affect behaviors. It also allows measuring certain personal characteristics, e.g., risk aversion level, that might be important to explaining behaviors but hard to measure outside laboratory. For these and other reasons, randomized experiments in laboratory are not subject to various limitations of observational experiments (Rosenbaum (2002)). Randomized experiment therefore offers an excellent methodology for us to answer the research question without worrying about confounding effects that might arise from using archival data.

Consistent with IRS's practice of disclosing only broad statistics like the audit rate of each income class, prior experimental studies on tax compliance usually consider settings where subjects are told to be audited independently at a known, constant probability (e.g., Allingham and Sandmo (1972), Yitzhaki (1974), Moser et al. (1995), Zimbelman and Waller (1999), Boylan and Sprinkle (2001), Kim et al. (2005), Kim and Waller (2005), Alm et al. (2009), and Kleven et al. (2011)). A recent theoretical study shows that such a *flat-rate* audit rule in equilibrium has the same deterrence effect as a variable-rate rule, referred to as the *bounded* rule (Yim (2009)).

Simply put, the bounded rule fully utilizes a given audit capacity to randomly select a sample of equally suspicious reports to check if the number of such reports exceeds the capacity, or otherwise audits all of such reports. Because the number of reports selected for audit is bounded by the audit capacity, the audit probability facing a taxpayer varies depending on the total number of suspicious reports filed by the taxpayer population. By setting the audit capacity appropriately, the compliance level induced by the bounded rule can be equivalent to that by the flat-rate rule. This theoretical equivalence together with the simple binary-income setting from which the bounded rule was derived makes comparing the two rules experimentally using human subjects a suitable way to answer our research question.

In designing our experiment, we bear in mind the "Why People Pay Tax" (WPPT) puzzle documented in the tax compliance literature (Alm et al. (1992)). It is unclear why most people file tax returns honestly when the average audit probability is only 1% (Slemrod (2007)). Given this phenomenon, it is important to ensure that our experiment provides a sufficient incentive to lie. Otherwise, if nearly all participants behave honestly in our experiment, the result would bias toward a "no difference" conclusion. To avoid this bias, the baseline *Flat-rate* treatment of our experiment provides a strong enough incentive for participants to lie. It is so strong that theoretically all participants should lie, just like the key feature of the documented puzzle. However, also like the puzzle, the actual outcome is a compliance level much higher than 0%.

To compare with the *Flat-Rate* treatment that represents the practice of disclosing only the average audit probability, our *Bounded* treatment lets participants know how the audit probability is contingent on the total number of suspicious reports filed by taxpayers. Like the *Flat-rate* treatment, the *Bounded* treatment has a predicted compliance level of 0%. The actual outcome, again, is far above the theoretical prediction.

We find that compared to the *Flat-rate* treatment, the compliance level is not lower under full disclosure of the contingent audit probability structure in the *Bounded* treatment. Interestingly, it is actually higher in estimated magnitude (43%, rather than 39% in the *Flat-Rate* treatment), though not statistically significant. Similar results continue to hold when confining to the last 10 periods where participants should have become familiar with the environment (47%, rather than nearly 42% in *Flat-Rate*). The findings support our hypothesis that there is no difference in the compliance levels under the bounded and flat-rate rules, which represent disclosing audit policy details (i.e., the contingent audit probability structure) versus merely the average audit probability. We conclude that disclosing audit policy details does not necessarily reduce tax compliance.

To see whether the conclusion might be sensitive to a parameter in the experiment, we contrast the *Bounded* treatment with the *Bounded*-hi-q treatment. This additional treatment captures the case where taxpayers in an area under the jurisdiction of an IRS District Office are more likely to have a high income.[5]. When the parameter q is high, there are multiple equilibria under the bounded rule in the tax compliance game of the experiment. One of the

[5] Consistent with the emphasis by Yim (2009), tax audits are administrated by IRS District Offices under audit capacity constraints. See further discussion in section 5.

equilibria involves all high-income taxpayers tacitly coordinating to lie. Another involves all of them reporting honestly. The third is a mixed equilibrium where each taxpayer randomizes to lie with the same probability.

We find that the conclusion from the first two treatments is not sensitive to the existence of multiple equilibria in the *Bounded*-hi-q treatment with a high q. We observe a higher level of compliance in this treatment (66% overall and 74% for the last 10 periods) than in both the *Flat-rate* and the *Bounded* treatment. We further conclude that fully disclosing the audit probability structure, rather than merely the average audit probability, can increase tax compliance, instead of reducing it.

Besides the main findings above, we also analyze the audit budget implications of the bounded rule to see whether they are broadly consistent with the theoretical insights of Yim (2009) where the rule was derived. The results suggest that the bounded rule on average conducts fewer audits than the flat-rate rule. If taking into account the budget commitment required to credibly implement the flat-rate rule, the bounded rule has a higher budget usage ratio than the flat-rate rule. Both results are in line with the theoretical insights about the bounded rule, suggesting no unexplained issue that might cause any concern.

Though consistent with the documented WPPT puzzle, the compliance levels observed in the experiment are under-predicted quite substantially by the standard theory. This leads us to conduct additional analyses to reconcile the discrepancy using alternative choice models under uncertainty. The observed behaviors can be satisfactorily explained by a loss aversion model. We are not aware of any unusual results from the analyses that might compromise the conclusion of our main analysis.

This paper adds to the literature on understanding how dissemination of enforcement information might affect taxpayers' behavior (e.g., Slemrod et al. (2001)). Focusing on the compliance impact of information dissemination regarding audit results, Alm et al. (2009) find that the effect of post-audit information is conditional on whether the taxpayer is well informed of the audit probability prior to filing. Unlike them, we do not consider disclosing population-wide audit results of the previous period before the filing in a period. Instead, to be in line with the setting of Yim (2009), we use each period to capture a new observation of a one-shot game, rather than a snapshot of a multiple-period dynamic game. Our focus is on the disclosure of the underlying contingent audit probability structure (*Bounded*), which has a deterrence effect theoretically equivalent to that of the average audit probability disclosed (*Flat-rate*). Alm et al. (2009) consider only the latter case to contrast with the alternative of no disclosure at all.

Research in the disclosure literature has predominantly concentrated on corporate transparency (e.g., Bushman et al. (2004) and Francis et al. (2009)). One of the main themes is that companies with more disclosure might enjoy the benefits from reducing information asymmetry, namely a lower cost of capital, a smaller bid-ask spread, etc (e.g., Botosan and Plumlee (2002)). This paper extends the literature to consider government transparency. What motivates government agencies' lack of transparency appears to be the potential benefits from being opaque. Our findings, however, suggest that a presumed benefit might not exist. Interestingly, there might even be some overlooked cost (in terms of foregone benefit) under certain circumstances (e.g., a high q).

Findings from accounting research suggest that investors do not fully exploit publicly available information, nor fully understand the implications of the information, in making investment decisions (e.g., Bartov et al. (2000), Dechow et al. (2008), and Landsman et al. (2011)). Possible reasons include limited attention or other information processing or transaction costs (e.g., Hirshleifer and Teoh (2003), Louis and Sun (2010), and Corwin and Coughenour (2008); see also the discussion by Schipper (2007)). Consistent with such findings, our results suggest that maybe IRS has overly worried about the impact of disclosing audit policy details.

To meet people's increasing demand for transparency in government, IRS can set out a plan to disclose more information about the audit policy on an annual basis. Each year the incremental disclosure should be about a clearly defined set of new information and be released on a specific date before the deadline of another round of tax return filing. This way researchers can precisely analyze the impact of the incremental information disclosure. Further evidence can thus be provided to determine whether even more disclosure or IRS's current position of information withholding should be supported. Ultimately, such research may help IRS to understand how its commissioned objectives can be best fulfilled. The unintended monitoring functions of IRS on the financial market and financial reporting quality (see Hanlon and Heitzman (2010), page 138, El Ghoul et al. (2011), and Hanlon et al. (2011)) might also be enhanced.

The organization of this paper is as follows. We describe the experiment design and procedure in the next section, ending with our hypothesis for testing. Main results from the experiment are discussed in section 3. In section 4, we conduct additional analyses to reconcile the discrepancy between the compliance level actually observed in the experiment and that predicted by the standard theory. Section 5 reviews related tax compliance studies and further explains why we design the experiment based on the bounded rule. Section 6 contains concluding remarks. The theoretical analysis upon which our experimental study is based, technical details and proofs, and the experiment instructions are provided in the appendix (Tan and Yim (2011)).

2 Experiment and Hypothesis

2.1 Design

The tax compliance game in all treatments of our experiment has three stages: (i) income reporting and tax deduction, (ii) audit and fine deduction, and (iii) feedback. Subjects receive either a high income $I_{\mathrm{H}} =$ €25 (H-type) or a low income $I_{\mathrm{L}} =$ €10 (L-type) with probability q or $1 - q$, respectively. Subjects are informed of the group size N and the probability q. Based on the capacity constraint in the lab, the size of the taxpayer population is fixed to be $N = 8$. The parameter q is either 0.5 or 0.9 depending on the treatment.

During the income reporting stage, subjects have to decide simultaneously and independently the report type ("high income" or "low income") to submit to an auditor, which is simulated by a computer. The computer automatically deducts taxes according to the reported income. The tax for subjects reporting a "high income" is $T_{\mathrm{H}} =$ €12.5, whereas the tax for subjects reporting a "low income" is $T_{\mathrm{L}} =$ €2.5.[6] Subjects are told that taxes

[6]Experimental parameters concerning taxation are chosen to be in line with the reality. For instance, the real-world tax rates for high-income and low-income taxpayers are usually dependent on the levels of their incomes. In particular, many countries such as Britain, the Netherlands, Germany, Italy and the USA use a progressive tax system instead of a proportional one. Hence, this experiment adopts a progressive tax system for the sake of facilitating subjects' understanding.

are deducted based on their reported income instead of true income. For instance, if H-type players submit a "low-income" report, they receive €22.5, instead of €12.5. Similarly, L-type players receive −€2.5, instead of €7.5, if they submit a "high-income" report.[7] In the audit stage, the computer implements either a flat-rate rule or a bounded rule to audit "low-income" reports submitted. In the experiment, "high-income" reports are not audited. This is consistent with the IRM guidelines (see section 5 for further details).

Described below are the designs of the three treatments of the experiment. Key parameters of the treatments are summarized in Table 1.

Flat-rate: In this treatment, subjects are told that those filing "low-income" reports independently face an audit probability of $a = 0.4$. This audit probability induces the same compliance rate as the bounded rule does with an audit capacity $K = 2$.[8] If subjects report honestly, nothing will happen to their final payoffs. If cheaters are caught by the auditor, they need to pay back the €10 of taxes evaded plus a fine of $F =$ €10.

Bounded: In this treatment, the fine for cheaters is exactly the same as in the *Flat-rate* treatment. The audit probability, however, depends on the total number of "low-income" reports received. The maximum number of audits to be conducted is $K = 2$. This value of the parameter guarantees a unique Nash equilibrium based on non-cooperative game theory (see the theoretical analysis provided in the appendix for details). Setting $K = 2$ means that if the number of "low-income" reports does not exceed two, then all of them will be audited with probability 1. Otherwise, the audit probability decreases monotonically with the number of "low-income" reports, denoted by L. In particular, the probability is 0.67 for $L = 3$; 0.5 for $L = 4$; 0.4 for $L = 5$; 0.33 for $L = 6$; 0.29 for $L = 7$; 0.25 for $L = 8$. Instead of merely disclosing the average audit probability, the contingent audit probability structure is fully disclosed to subjects through an audit probability table (see the experiment instructions provided in the appendix for details).

Bounded-hi-*q:* Except for the ex-ante probability q of receiving a high income, this

[7] Even when a subject with a low income makes a loss by submitting a "high-income" report and that decision is selected for payment, the potential loss is covered by a show-up fee of €3. During the experiment sessions, this situation never actually happens.

[8] Because the flat-rate rule induces all-or-none behavior in compliance, such a rule with an audit probability $a < 0.5$ theoretically has the same deterrence effect as the bounded rule, assuming the standard setup with perfectly rational, risk-neutral players.

treatment is the same as the *Bounded* treatment. The high $q = 0.9$ of this treatment represents the case of an area under the jurisdiction of an IRS District Office where taxpayers are more likely to have a high income. Compared to the *Bounded* treatment, subjects lying in this treatment face a higher degree of uncertainty because fewer “low-income” reports will be submitted given the low probability of having low-income taxpayers. Consequently, there will be fewer honest “low-income” reports to pool with lying “low-income” reports, making lying easier to be detected by audits. The theoretical analysis provided in the appendix shows that the game in this treatment has multiple equilibria. We are interested in knowing whether the behavior observed in the *Bounded* treatment is sensitive to the presence of multiple equilibria under the bounded rule when q is high.

2.2 Procedure

The experiment was conducted at the laboratory of a European university from October to December 2009. Most of the university students participating as subjects in the experiment were major in economics or business. The experiment instructions, provided in Appendix C.2, were modified from those in prior tax compliance studies, namely Alm et al. (2009), Kim et al. (2005), and Kim and Waller (2005). We used Z-Tree software (Fischbacher (2007)) to program and conduct the experiment.

Each treatment of the experiment consists of four sessions; each session has 16 subjects. The duration of a session is about 1 hour (including the initial instruction and final payment to subjects). The average earnings are €16.23 (including the €3 show-up fee). At the beginning of each session, subjects are randomly assigned to the computer terminals. Before the experiment starts, subjects have to complete an exercise making sure that they understand the rules of the tax compliance game.

The game consists of 30 periods. At the beginning of each period, 16 subjects are randomly allocated into two groups of eight. The random re-matching protocol minimizes the chances that subjects encounter the same group of participants again. The purpose is to simulate a one-shot scenario but allows the subjects to be familiar with the game environment. This is particularly important for treatment sessions with the bounded rule. Each period can thus constitute a new observation of a one-shot game, rather than a snapshot of a multiple-period dynamic game. At the end of each period, a summary screen is presented to subjects with feedback information including the subject's true and reported income, and the final payoff for the period. Subjects are not informed of others' payoffs.

Upon finishing the tax compliance game part of the experiment, subjects are asked to complete a risk elicitation task similar to the one used by Holt and Laury (2002). The instructions for the risk elicitation task are handed out only after the tax compliance game. Hence, the subjects are not aware of its existence beforehand. Details of this task can be found in the experiment instructions. The task measures subjects' risk aversion levels, which could be useful in explaining their behaviors.

During the payment stage, one period of the tax compliance game and the realization of one lottery of the risk elicitation task are randomly selected to determine the final payment to a subject. This random payment scheme mitigates the potential income effect that the subjects carry across different periods of the game and over to the risk elicitation task.

We conclude this section by stating the hypothesis for testing, which is based on the prediction (Proposition 2) derived in the theoretical analysis given in the appendix.

Hypothesis 1 *The underreporting rates in the Flat-rate and Bounded treatments are the same.*

Because the game in the *Bounded*-hi-q treatment has multiple equilibria, we merely compare the underreporting rate in the treatment with those in the other two without advancing any hypotheses based on theoretical predictions.

3 Main Results

Figure 1 depicts the average underreporting rates across treatments. The dynamics in the *Flat-rate* and *Bounded* treatments look similar. In contrast, the average underreporting rate in the *Bounded*-hi-q is visibly lower and declines steadily over periods.

Table 2 summarizes the compliance behaviors and auditing statistics across experimental treatments. The first three columns contain averages over all 30 periods of play. The next three columns are averages of the last 10 periods, where subjects' behaviors are expected to be more stable after becoming familiar with the environment. Statistical testing on the treatment effects is based on the two-sided Wilcoxon rank-sum test (also called the Mann-Whitney test). We adopt the strictest standard to use each session as an independent observation. This avoids any doubt that observations at more refined levels (e.g., by subject or by session-period) might not be completely independent. Such doubt arises from the fact that unlike individual decision-making experiments, subjects in our treatments under the bounded rule interact with each other, rather than make their own independent decisions; moreover, their behaviors might be correlated across periods.[9]

We first focus on the *Flat-rate* and *Bounded* treatments. The top panel of the table reports statistics concerning all subjects. The first row of the panel indicates that the actual frequency of being an H-type in the two treatments is very close to the pre-specified levels. The second row displays the percentage of "low-income" reports out of all reports received (i.e., the total number of reports from L-type players or lying H-type players, divided by 8). The ratio is around 80% in the two treatments.

The middle and the bottom panels of the table provide data for testing our hypothesis and examining the audit budget implications of the experiment results. Our findings are summarized as follows:

Result 1 *Hypothesis 1 is supported. The difference between the underreporting rates observed in the Flat-rate and Bounded treatments is statistically insignificant.*

Support: The average underreporting rate is 60.83% in the *Flat-rate* treatment and 57.11% in the *Bounded* treatment. A two-sided Wilcoxon rank-sum test cannot reject the null hypothesis that the underreporting rates of the two treatments are the same ($p = 0.386$). In

[9] To see whether our conclusion on Hypothesis 1 is robust, we also have done a two-sample z-test, with 64 observations (at the subject level) for the *Flat-rate* treatment and 4 observations (at the session level) for the *Bounded* treatments. The hypothesis is still supported. In addition, we have done the test with 64 observations (at the subject level) for each of the *Flat-rate* and *Bounded* treatments. Even with this liberal interpretation of the independence of the observations from the *Bounded* treatment, the test result continues to support the hypothesis. Finally, we have run a logistic regression using subject-period observations. The dependent variable equals 1 if an H-type subject underreports in a period and 0 if s/he honestly reports in the period. The independent variable is whether the observation comes from the *Bounded* or *Flat-rate* treatment (with or without social demographic controls). Based on a standard error corrected for clustering by subject, the estimated coefficient of the treatment variable is statistically insignificant at the 5% level. This further confirms that the underreporting rates in the two treatments are statistically indistinguishable.

the last 10 periods, the magnitude of the difference in underreporting rate becomes slightly larger but still statistically insignificant ($p = 0.564$).

Result 2 *The bounded rule is more cost-effective in the sense that on average*
(i) fewer audits are performed, and
(ii) the budget-usage ratio is higher
in the Bounded treatment than in the Flat-rate treatment.

Support: The difference in the *cheater detection rate*, namely the frequency that a tax cheater is caught during an audit stage, is not statistically significant in both treatments ($p = 0.113$ for all 30 periods; $p = 0.149$ for the last 10 periods). This means that the bounded and flat-rate rules are equally effectively in detecting cheaters.

Several pieces of evidence support that the bounded rule is more cost-effective. Assuming a constant cost per audit, we can use the number of audits performed in a treatment as a proxy for the audit resources consumed to achieve the compliance level observed. Both the *total and the average number of audits* performed are significantly lower in the *Bounded* treatment ($p < 0.05$).

We also look at the *audit selection rate*, which is defined as the proportion of "low-income" reports selected for audit, out of the total number of such reports received. This rate is significantly lower in the *Bounded* treatment, both for all 30 periods or only the last 10 periods ($p < 0.05$). These results suggest that auditing with the bounded rule can achieve the same compliance level at a lower cost.

Finally, we look at the *budget usage ratio*, which is defined as the percentage of audit resources actually used, out of the budget commitment required to credibly support an audit rule. The ratio is 100% in the *Bounded* treatment, which means that all resources committed are used at the full capacity in each period (i.e., two audits). Under the flat-rate rule, the budget-usage ratio is only 32%. The inefficiency is due to the fact that in order to credibly implement the flat-rate rule, the auditor must have the resources to be ready to do all eight audits in each period. However, much fewer audits are actually carried out.

In an equilibrium setting, Yim (2009) has analytically shown that even when the flat-rate rule can be implemented using large-sample random sampling, the budget usage ratio remains substantially below that of the bounded rule. Unfortunately, we cannot assess this theoretical insight with our experiment because the size of the experimental taxpayer population is only eight subjects.

The following is our result from the *Bounded*-hi-q treatment.

Result 3 *The underreporting rate is significantly lower in the Bounded-hi-q treatment than in the Bounded and Flat-rate treatments. The higher compliance level is achieved with significantly fewer audits performed and with a higher budget-usage ratio.*

Support: The average underreporting rate in the *Bounded*-hi-q treatment is 33.95% over all 30 periods and 26.16% in the last 10 periods. The compliance level in this treatment is the highest, as the underreporting rate is significantly lower compared to the other two treatments ($p < 0.05$). The difference is already salient in the first period and remains highly significant throughout the other periods of the game.

Regarding auditing statistics, the total number of audits performed is smaller in this treatment than in the *Bounded* treatment ($p < 0.05$). However, the audit selection rate turns out to be significantly higher ($p < 0.05$), owing to fewer "low-income" reports received given the higher q in this treatment. The cheater detection rate is remarkably higher as well ($p < 0.05$). The budget-usage ratio is 95.63%, which is significantly higher than that in the *Flat-rate* treatment (32.03%).

4 Additional Analyses

While the main results discussed above have answered our research question concerning the impact of disclosing audit policy details, the observed compliance levels remain unexplained by the standard theory. In this section, we make an attempt to better understand individual-level compliance behavior. The purpose is to ensure that we have not overlooked anything that might lead to misinterpretation of the main results.

4.1 Stochastic Nature of Individual-level Behavior

Figure 2 displays the frequency distributions of the individual underreporting rate across treatments. The horizontal axis represents a subject's *individual underreporting rate*, i.e., the percentage of times where the subject when assigned as a high-income taxpayer submits a "low-income" report. The vertical axis represents the percentage of subjects with similar underreporting rates in a treatment.

The main message conveyed by Figure 2 is that the standard theory has limited explanatory power over the individual-level data of the *Flat-rate* and *Bounded* treatments. Only 29.13% of the subjects in the *Flat-rate* treatment and 23.43% of those in the *Bounded* treatment underreport whenever receiving a high income, behaving in accordance with the standard theory. The percentage of seemingly intrinsically honest subjects, who always report their income truthfully, is 12.5% in the *Flat-rate* treatment and 15.63% in *Bounded*. Even after correcting for the presence of seemingly intrinsically honest players, the standard theory still underpredicts the compliance levels observed in the treatments.

Figure 2 also indicates that around 60 percent of the subjects switch between the two options at various levels of frequency. This pattern is very similar in the two treatments (Mann-Whitney test: $p = 0.322$). In contrast, the distribution of the underreporting rate in *Bounded*-hi-q is significantly different (Mann-Whitney test: $p < 0.05$). Throughout this treatment, only about 7% of the H-type choose to submit "low-income" reports, whereas 33% of them honestly report a "high income."

Recognizing the highly stochastic nature of individual-level behavior, we conduct additional analyses to better understand the behavior using several choice models under uncertainty. Because the game in *Bounded*-hi-q has multiple equilibria and the compliance behavior observed in the treatment appears to follow a different pattern, we focus on the *Flat-rate* and *Bounded* treatments in our attempt to explain the stochastic component of the behavior.

4.2 Choice Models under Uncertainty

The standard theory predicts that strategic players will always choose to submit "low-income" reports. On the other hand, intrinsically honest players will always report the type of income they receive. In either case, the choices should be consistent across periods. In contrast, Figure 2 suggests that many participants in our experiment make stochastic choices, which is consistent with McFadden's discrete-choice framework (McFadden (2001)).

This framework relaxes the perfect rationality assumption to accommodate boundedly rational behavior. Models in this framework are motivated by empirical studies where observed decisions exhibit some noise (see, e.g., Fischbacher and Stefani (2007), Loomes (2005), Rieskamp (2008), and Wilcox (2011)). Such noise could come from observed sources like decision errors. It could also arise from unobserved or unmodeled channels such as individual perceptions of the game or sensitivity to payoff changes. The presence of such noise leads to people making decision errors and hence behaving inconsistently with their choices.

The *Flat-rate* treatment is essentially a non-strategic choice-under-uncertainty problem for H-type players. Therefore, the classic individual discrete-choice model is a natural choice to explain the stochastic individual behavior. The *Bounded* treatment introduces interactions among subjects. A general way to incorporate decision errors into a strategic interaction setting is the quantal response equilibrium first proposed by McKelvey and Palfrey (1995). This equilibrium concept is based on McFadden (1973)'s random utility maximization model of the same framework.

Using the discrete-choice framework, we estimate and compare three choice models under uncertainty. They are risk aversion, loss aversion, and loss aversion with probability weighting. Brief descriptions of the models follow. (See the appendix for further details of the loss aversion model and of the discrete-choice framework applied to our experimental setting.)

Risk aversion. The first model we consider simply relaxes the assumption of risk neutrality. In the risk aversion model, subjects are assumed to have a constant relative risk aversion (CRRA) utility function: $u(\pi) = (\pi^{1-r}) \,/\, (1-r)$, where π is the disposable income (i.e., after-tax income) and r is the CRRA coefficient. This model offers the possibility of explicitly testing the assumption of risk neutrality. If the estimated r is significantly different from zero, then the null hypothesis that subjects are risk neutral can be rejected. We have also considered alternative utility forms such as constant absolute risk aversion (CARA) and power-expo. There is little change in the goodness of fit to the data.

Loss aversion. While the observed compliance behavior can be explained by risk attitude, it is also consistent with the notion of loss aversion. Recent research has shown that loss aversion provides a much better account of tax evasion both in the lab and in the field (see, e.g., Elffers and Hessing (1997), Yaniv (1999), King and Sheffrin (2002), and Dhami and Al Nowaihi (2007, 2010)). The loss aversion model characterizes individuals as loss-averse in terms of the disposable income relative to some reference income. For a given amount of such relative income $x > 0$ and a value function $v(x)$, losses are weighted more than gains, i.e., $|-v(-x)| > v(x)$. We consider Tversky and Kahneman (1992)'s specification of the value function: $v(x) = x^{\alpha}$ if $x \geq 0$, and $v(x) = -\lambda(-x)^{\beta}$ if $x < 0$. The α and β are the parameters determining the curvature of the function, and λ is the coefficient of loss aversion. Subjects are considered loss-averse if $\lambda > 1$.[10]

Loss aversion with probability weighting. Besides the value function, subjects could also have a nonlinear transformation of the probability scale. For example, Kahneman and Tversky (1979) find that people overestimate low probabilities and underestimate high probabilities. To examine the effect of subjective probability weighting, we also estimate a model of loss aversion with probability weighting. In particular, we consider a popular form of the one-parameter probability-weighting function: $w(\gamma) = \gamma^{\delta}/(\gamma^{\delta} + (1-\gamma)^{\delta})$, where γ is a probability and $\delta \geq 0$ is the weighting parameter. Note that if $\delta < 1$, the weighting function has an inverted-S shape, which is concave for low probabilities and convex for high probabilities, and crosses the diagonal at the probability of 1/3.

Effectively speaking, H-type players' reporting decision is like choosing between a safe option (honest reporting) and a risky lottery (underreporting), with known, constant probabilities in the *Flat-rate* treatment but unknown, endogenous probabilities in the *Bounded* treatment. Thus, the reporting choice in the *Bounded* treatment is affected by the subjects'

[10] Given a fixed payoff structure, data from the tax compliance game alone contain only two moments (i.e., the percentages of subjects selecting the "risky" lottery in the *Flat-rate* and *Bounded* treatments). They lack sufficient identification power to estimate three parameters jointly. Therefore, we pool together the data from the risk elicitation task and the tax compliance game to jointly estimate the parameters.

perceived average audit probability, denoted by $\widehat{a}$. Our analyses let us infer an estimate of $\widehat{a}$. With the estimate, we can answer the following questions: What average audit probability of a flat-rate rule would induce the same level of compliance as observed in the *Bounded* treatment? Moreover, how do risk aversion, loss aversion, and probability weighting influence the subjects' perception of the average audit probability in the *Bounded* treatment?

4.3 Additional Result

Table 3 reports the estimation results of the three models based on the *Flat-rate* and *Bounded* treatments. All coefficient estimates of the models are highly significant (at the 1% level), suggesting that all of them are useful in explaining the compliance levels observed in the treatments. For instance, the risk aversion specification suggests that subjects are risk averse in both treatments, as the CRRA coefficient r is significantly larger than zero. It indicates that risk aversion helps in explaining the data. The perceived audit probability in the *Bounded* treatment is 0.336. In other words, a flat-rate rule with an audit probability of 0.336, rather than 0.4, would induce such risk-averse subjects to comply at a level similar to what has been observed in the *Bounded* treatment.

Results of the loss aversion specification suggest that subjects are loss-averse. The estimated coefficient of loss aversion λ is larger than 1 in both treatments, which means that subjects are more sensitive to a loss than a gain of the same magnitude. The estimated slope coefficients of the value function indicate concavity in the gain domain (α) and convexity in the loss domain (β). Moreover, a Vuong test on non-nested models favors the loss aversion model over the risk aversion model ($p < 0.05$). For loss-averse subjects, a flat-rate rule with an audit probability of 0.306 would induce the same compliance level as observed in the *Bounded* treatment.

The third specification combines loss aversion with probability weighting. We find that this specification does not improve the goodness of fit significantly. Moreover, the probability-weighting parameter δ is not significantly different from 1 for both treatments ($p = 0.438$ and 0.397 for *Flat-rate* and *Bounded*, respectively). So the subjectively weighted probabilities used by the subjects on average are in line with the objective probabilities. Overall, the results suggest that what drives the observed compliance level is likely to be the way the subjects view losses and gains, rather than how they assess probabilities.

Figure 3 displays the predicted underreporting rates based on different models and the actual rates observed in the treatments. Because probability weighting adds little to the loss aversion model, the predicted underreporting rate of this model is based on coefficients estimated without probability weighting. The compliance behaviors in our experiment are best explained by the loss aversion model, compared to the alternatives, namely the risk aversion model and the standard theory with perfectly rational, risk-neutral players (with and without correction for seemingly intrinsically honest subjects). We conclude this section with the following result.

Result 4 *The compliance levels observed in the Flat-rate and Bounded treatments can be satisfactorily explained by a loss aversion model under the discrete-choice framework.*

The additional analyses in this section solve the otherwise unexplained levels of compliance observed in the experiment. Throughout the process, we do not find anything that might compromise the conclusion from our main analysis.

5 Relation to Tax Compliance Literature

IRS has the practice of disclosing only broad statistics such as the audit rate of each income class. In line with this, many tax compliance studies consider settings where taxpayers are told to be audited independently at a known, constant probability (see literature reviews by Andreoni et al. (1998), Alm and McKee (1998), and Slemrod and Yitzhaki (2002)). Such a setting is captured by the *Flat-rate* treatment in our experiment.

In the vast majority of tax compliance studies, the attention is on the interaction between the auditor and a taxpayer, without considering the interaction with the whole taxpayer population, or the interactions among taxpayers. A notable exception is Alm and McKee (2004), who experimentally study a "DIF" rule that represents IRS's audit policy based on discriminant function (DIF) scores. The audit probability of their "DIF" rule depends on the deviation of an individual's reported income from the average of the incomes reported by all other players. This audit rule induces a coordination problem for taxpayers who want to cheat on taxes. In their experiment, all participants receive the same level of income in any given period. This is not the case in our experiment. Besides this distinction leading to a different coordination problem in the *Bounded* treatment, another difference is that the interaction induced by the bounded rule among taxpayers does not always lead to a coordination game.

Tax compliance studies rarely explicitly consider audit budget. Unlike others, Yim (2009) emphasizes the importance of the budget commitment required to support an audit policy and the implication to the structure of the policy. Using a setting similar to the classic tax compliance game (Graetz et al. (1986)), he shows that the equilibrium audit policy that minimizes the required committed budget takes the form of the bounded rule. Such a binary-income setting, or similar discrete-type extensions, have been used in many studies (e.g., Mills et al. (2010), Mills and Sansing (2000), and some others cited in footnote 4 of Yim (2009)).

Though stylized, the binary-income setting captures some salient features of audit selection in reality. For example, low-income taxpayers in the setting have no incentive to submit "high-income" reports. So these reports must have been submitted by high-income taxpayers. Because *auditing such reports cannot lead to higher tax revenue, these reports are not audited* under either of the audit rules considered in our experiment. Indeed, the IRM prescribes that "[c]lassifiers [who review computer-prescreened tax returns to determine which are to be put forth for examination (i.e., audit)] should compare the potential benefits to be derived from examining a return to the resources required to perform the examination. Although you may identify some potentially good issues on the return, *if they would not yield a significant adjustment, the return should be accepted as filed.*" (emphasis added) (see paragraph 1 of IRM 4.1.5.1.5.1.1 (10-24-2006) in Section 5 "Classification and Case Building" of the manual). In line with this, a recent study by Phillips (2010) shows that IRS focuses on auditing taxpayers expected to have high unmatched income (i.e., income cannot be cross-checked with third-party reports such as Form W-2) and rarely examines taxpayers likely to have only matched income.

Besides Yim (2009), Erard and Feinstein (1994) also explicitly consider audit budget. However, like other tax compliance studies, they focus on the interaction between the auditor and an atomic taxpayer in the population. This effectively reduces the whole taxpayer population into a representative taxpayer. The complexity of the model gives rise to the characterization of the equilibrium by a second-order differential equation The equation does not have a closed-form solution and hence can only be solved numerically. In contrast, the setting of the classic tax compliance game is much simpler. Moreover, the bounded rule that constitutes an equilibrium audit strategy has a simple structure determined by the audit capacity constraint.

Indeed, audit capacity is an important concern in IRS's operations. Guidelines in the IRM suggest that a substantial part of the agency's operations is done at the District Office level, referred to as "[geographical] Area" in the manual. The audit capacity of each District Office, namely the staff force constituting mainly of revenue agents, tax compliance officers, return classifiers, etc (referred to as "posts-of-duty (POD)" in the manual), is determined based on the approved national examination plan constrained by resources requested in the Congressional Budget (see IRM 4.1.1.2 (10-24-2006) "Examination Plan").

Besides audit capacity, the bounded rule or the binary-income setting has other stylized features resembling IRS's audit policy. To point out the similarity, it is useful to begin with a quick overview of the audit selection procedure in reality. According to the IRM, tax returns are first computer-scored using the DIF System (see IRM 4.1.3.2 (10-24-2006) "DIF Overview"). Then with the national minimum cutoff score determined by National Headquarters each year, returns above the cutoff are added to the DIF inventory (see IRM 4.1.1.3 (10-24-2006) "Minimum DIF Cutoff Score").

Alm and McKee (2004) have studied a "DIF" rule that triggers an audit to a taxpayer based on the "deviation between his or her reported income and the average reported income" in an experiment session. Anecdotal evidence, however, indicates that what matters most is not the reported income of a return relative to others' average. "[T]ax professionals, who are familiar with I.R.S. procedures, say that the [DIF] formula examines the relationships between those income and deduction items that the I.R.S. has found to be the best indicators of compliance, chiseling and cheating." (Johnston (1996)). In line with this, a statistics professor Aczel (1994) has used a "supercomputer and modern statistical techniques like logistic regression or classification and regression trees to determine which kinds of returns get audited" (Johnston (1996)) and found that "taxpayers whose Schedule A deductions are less than 35 percent of income are almost never audited, while those who deduct 44 percent or more of income are almost certain to be audited. Those who fall in between those figures are at risk of being audited, depending on which type of deductions they take." (Johnston (1995a)).

Thus, whether certain deduction items have been claimed and their amounts relative to the reported income of the return seems to be most important. A return would have little chance to be added to the DIF inventory if "suspicious" deduction items were not claimed. The red-flag nature of claiming "suspicious" deduction items is similar to the pooling of "low-income" reports by lying taxpayers with those by honest low-income taxpayers in the binary-income setting of the experiment.

Not every return added to the central DIF inventory will eventually be audited. To be selected for audit, a return must first be among those ordered by a relevant Area for classification into *accepted as filed* or *selected for examination* (i.e., audit) (see IRM 4.1.5.1.3 (10-24-2006) "Sorting of Classified Returns"). Areas might have different selection rates for a variety of reasons (e.g., local issues, classifiers' judgment, etc). Therefore, to meet the audit target of an Area in the Examination Plan, "[t]he PSP [(i.e., Planning and Special Programs Territory Manager)] will calculate the Area DIF cutoff score ... giving consideration to the selection rate." (see IRM 4.1.1.3.1 (10-24-2006) "DIF Cutoff Score"). With the Area DIF cutoff, returns of an Area are divided into two groups: *above-cutoff* returns (analogous to the "low-income" reports in the experiment) and *below-cutoff* (analogous to the "high-income" reports).

Areas order returns from the central DIF inventory based on their specific cutoffs. After classification, returns selected for audit are categorized into "Field Examination" (i.e., visits at taxpayers' sites) or "Office Examination" (i.e., interviews at IRS offices) (see IRM 4.1.5.1.3 (10-24-2006) "Sorting of Classified Returns"). The returns are added to the Examination inventory (see IRM 4.1.1.6.3 (10-24-2006) "Inventory Monitoring"). Later the audits of these returns are assigned to POD's (i.e., revenue agents, tax compliance officers, etc) "based on ZIP codes [on the returns] using the ZIP/POD Lookup Table" (see IRM 4.1.1.7 (10-24-2006) "ZIP/POD Tables").

The IRM has guidelines to regulate the flow of orders in accordance with the Examination Plan. A POD Supplement Order is allowed as an exception if "there is a workload shortage at a specific POD" (see IRM 4.1.3.4 (10-24-2006) "Guidelines for Ordering Returns"). Nonetheless, the IRM specifies that if such orders "result in the delivery of returns that are below the [Area] DIF cutoff score", "not more than 10% of the returns ordered for any POD should be below the DIF cutoff score." (see IRM 4.1.1.3.2 (10-24-2006) "Use of DIF Cutoff Score for Return Orders"). In other words, aside from the 10% flexibility, a POD is not permitted to audit below-cutoff returns even when the POD has audited all the above-cutoff returns assigned to it, with idle capacity to audit more. This feature is similar to the key characteristic of the bounded rule: audit as many as possible if the number of suspicious reports exceeds the given capacity, or otherwise audit all such reports but none of the unsuspicious despite under-utilized capacity.

Because of the simple setting, the similarity with key features of the reality, and the theoretical equivalence to the flat-rate rule's deterrence effect, we use the bounded rule to represent the underlying audit policy of a tax authority that discloses merely the average audit probability to taxpayers.

6 Concluding Remarks

Tax authorities around the world often are reluctant to disclose audit policy details.[11] In particular, the US IRS has the practice of releasing broad statistics like the audit rate of each income class but opposes pressures demanding details on how different circumstances might result in a higher audit probability to taxpayers. In this paper, we ask whether the potential adverse impact on tax compliance could be a serious concern justifying the reluctance of tax authorities like IRS to disclose audit policy details.

To answer the question, we carefully consider the theoretical deterrence-equivalence of two audit rules and the documented WPPT puzzle in designing the treatments of our experiment. In the *Flat-rate* treatment, participants are told that they independently face a known audit

[11] "[M]ost [tax] agencies undertake substantial precautions to maintain the secrecy of their audit selection procedures." (Andreoni et al. (1998)). See OECD (2004) for an overview of various countries' audit selection systems.

probability. By contrast, participants in the *Bounded* treatment are fully informed of the contingent audit probability structure. We first show that according to the standard theory, participants should have a sufficiently strong incentive to lie about their income, regardless of the treatments. Based on this theoretical prediction that is consistent with the WPPT puzzle, we develop the hypothesis for testing.

Our findings show that consistent with the WPPT puzzle, the observed compliance levels are substantially higher than the theoretically predicted levels. Most important, the compliance levels of the two treatments that represent merely disclosing the average audit probability versus fully disclosing the audit policy details are not significantly different. This main result supports our hypothesis, suggesting that disclosing audit policy details do not necessarily reduce tax compliance. The examination with a third treatment to assess the sensitivity of our results to the existence of multiple equilibria suggests that disclosing audit policy details can increase, rather than reduce tax compliance.

We check two things to ensure that behaviors observed in the experiment are consistent with what we know from theories, and hence our main results are not compromised by anything that we could not explain. First, we verify that the audit budget implications of the observed behaviors are broadly consistent with the theoretical insights of the study where the equivalence between the bounded and flat-rate rules was derived. Then we use alternative choice models under uncertainty to explain the observed compliance levels under-predicted by the standard theory. Results from these exercises confirm what we know from theories. We therefore believe that our main result is not affected by some unknown factor.

Obviously, the evidence collected from one experiment cannot constitute a strong ground for tax authorities (sharing IRS's concern) to change their disclosure practices. Nevertheless, given the trend in increasingly stronger demand for government transparency, the evidence from this experiment does provide a reasonable basis for tax authorities to be more open-minded in viewing the issue. Compared to IRS, some agencies in other countries appear to be more liberal and transparent (see, e.g., Canada Revenue Agency and Australian Tax Office discussed in Hasseldine (2007) and Leviner (2008)). However, unless tax authorities let researchers examine more accurately and thoroughly the impacts of disclosing audit policy details, no one can tell what level of disclosure is best for society.

Let us re-iterate our suggestion already made in the introduction: IRS can set out a plan to disclose on a properly selected date of each year more information about the audit policy. This way researchers can precisely analyze the impact of the incremental information disclosure. Further evidence can thus be provided to determine whether even more disclosure or IRS's current position of information withholding should be supported.

References

Aczel, A. (1994). *How to Beat the I.R.S. at Its Own Game: Strategies to Avoid - and Fight - An Audit*. Four Walls Eight Windows.

Allingham, M. G. and A. Sandmo (1972). Income tax evasion: a theoretical analysis. *Journal of Public Economics 1*(3-4), 323–338.

Alm, J., B. R. Jackson, and M. McKee (2009). Getting the word out: Enforcement information dissemination and compliance behavior. *Journal of Public Economics 93*(3-4), 392–402.

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

Alm, J. and M. McKee (1998). Extending the lessons of laboratory experiments on tax compliance to managerial and decision economics. *Managerial and Decision Economics 19*(4-5), 259–275.

Alm, J. and M. McKee (2004). Tax compliance as a coordination game. *Journal of Economic Behavior & Organization 54*(3), 297–312.

Andreoni, J., B. Erard, and J. Feinstein (1998). Tax compliance. *Journal of Economic Literature 36*(2), 818–860.

Bartov, E., S. Radhakrishnan, and I. Krinsky (2000). Investor sophistication and patterns in stock returns after earnings announcements. *Accounting Review 75*(1), 43–63.

Botosan, C. A. and M. A. Plumlee (2002). A re-examination of disclosure level and the expected cost of equity capital. *Journal of Accounting Research 40*(1), 21–40.

Boylan, S. J. and G. B. Sprinkle (2001). Experimental evidence on the relation between tax rates and compliance: The effect of earned vs. endowed income. *Journal of the American Taxation Association 23*(1), 75–90.

Bushman, R. M., J. D. Piotroski, and A. J. Smith (2004). What determines corporate transparency? *Journal of Accounting Research 42*(2), 207–252.

Corwin, S. and J. Coughenour (2008). Limited attention and the allocation of effort in securities trading. *Journal of Finance 63*(6), 3031–3067.

Davis, S. L. (1997). *Unbridled Power: Inside the Secret Culture of the IRS.* New York: HarperCollins.

Dechow, P. M., S. A. Richardson, and R. G. Sloan (2008). The persistence and pricing of the cash component of earnings. *Journal of Accounting Research 46*(3), 537–566.

Department of the Treasury (2011). Chief FOIA Officer Report. March. http://www.treasury.gov/FOIA/Documents/Chief%20FOIA%20Officer%20Report%20 FINAL%20030711%20clean%20v2MD.pdf. Accessed on 17 May 2011.

Dhami, S. and A. Al-Nowaihi (2007). Why do people pay taxes? Prospect theory versus expected utility theory. *Journal of Economic Behavior & Organization 64*(1), 171–192.

Dhami, S. and A. Al-Nowaihi (2010). Optimal taxation in the presence of tax evasion: Expected utility versus prospect theory. *Journal of Economic Behavior & Organization 75*(2), 313–337.

El Ghoul, S., O. Guedhami, and J. Pittman (2011). The role of IRS monitoring in equity pricing in public firms. *Contemporary Accounting Research*, forthcoming. doi:10.1111/j.1911–3846.2011.01065.x.

Elffers, H. and D. J. Hessing (1997). Influencing the prospects of tax evasion. *Journal of Economic Psychology 18*(2-3), 289–304.

New York Times (1981a). Stevens Blocks Release Of Internal Revenue Tapes. March 6. http://www.nytimes.com/1981/03/06/us/around-the-nation-stevens-blocks-release-of-internal-revenue-tapes.html. Accessed on 17 May 2011.

New York Times (1981b). U.S. Asks Court to Nullify Order To Release Secrets on Tax Audits. May 14. http://www.nytimes.com/1981/05/14/us/us-asks-court-to-nullify-order-to-release-secrets-on-tax-audits.html. Accessed on 17 May 2011.

Time (1974). The Law: Blasting Facts Free. December 2. http://www.time.com/time/magazine/article/0,9171,945326,00.html. Accessed on 17 May 2011.

Erard, B. and J. S. Feinstein (1994). Honesty and evasion in the tax compliance game. *The RAND Journal of Economics 25*(1), 1–19.

Fischbacher, U. (2007). Z-tree: Zurich toolbox for readymade economic experiments. *Experimental Economics 10*(2), 171–178.

Fischbacher, U. and U. Stefani (2007). Strategic errors and audit quality: An experimental investigation. *Accounting Review 82*(3), 679.

Francis, J. R., S. Huang, I. K. Khurana, and R. Pereira (2009). Does corporate transparency contribute to efficient resource allocation? *Journal of Accounting Research 47*(4), 943–989.

Gates, B. E. (2000). Statement of Bryan E. Gates, Enrolled Agent and Member, Federal Taxation Committee, National Society of Accountants, Alexandria, Virginia. In *Hearing on the 2000 Tax Return Filing Season and the IRS Budget for Fiscal Year 2001, before the Subcommittee on Oversight of the Committee on Ways and Means of the 106th Congress House of Representatives*. Washington: US Government Printing Office.

Ginsberg, W. R. (2011). The Obama Administration's Open Government Initiative: Issues for Congress. *Congressional Research Service (CRS) Report for Congress*, January 28. http://www.fas.org/sgp/crs/secrecy/R41361.pdf. Accessed on 17 May 2011.

Graetz, M., J. Reinganum, and L. Wilde (1986). The tax compliance game: Toward an interactive theory of tax enforcement. *Journal of Law, Economics, and Organization 2*, 1–32.

Hanlon, M. and S. Heitzman (2010). A review of tax research. *Journal of Accounting and Economics 50*(2-3), 127–178.

Hanlon, M., J. L. Hoopes, and N. O. Shroff (2011). The effect of tax authority monitoring and enforcement on financial reporting quality. Available at SSRN: http://ssrn.com/abstract=1691158.

Hasseldine, J. (2007). Tax Implementation: Commentary. Prepared for the Report of a Commission on Reforming the Tax System for the 21st Century, Chaired by Sir James Mirrlees. The Institute for Fiscal Studies. http://www.ifs.org.uk/mirrleesreview/commentaries/hasseldine.pdf. Accessed on 22 May 2011.

Hirshleifer, D. and S. H. Teoh (2003). Limited attention, information disclosure, and financial reporting. *Journal of Accounting and Economics 36*(1-3), 337–386.

Holt, C. A. and S. K. Laury (2002). Risk aversion and incentive effects. *American Economic Review 92*(5), 1644–1655.

IRS (2010). IRS Processing Codes and Information (Formerly Titled: ADP and IDRS Information) - For Public Release. http://www.irs.gov/pub/irs-utl/document6209.pdf. Accessed on 5 May 2011.

Johnston, D. C. (1995a). I.R.S. Puts Off Plan For Detailed Audits Of Random Returns. *New York Times*, October 24. http://www.nytimes.com/1995/10/24/business/irs–puts–off–plan–for–detailed–audits–of–random–returns.html?pagewanted=print&src=pm. Accessed on 18 May 2011.

Johnston, D. C. (1995b). U.S. Archives Warns I.R.S. Over Secrecy. *New York Times*, December 21. http://www.nytimes.com/1995/12/21/business/us–archives–warns–irs–over–secrecy.html. Accessed on 17 May 2011.

Johnston, D. C. (1996). Some New Tricks To Help Filers Avoid An Old Audit Trap. *New York Times*, February 25. http://www.nytimes.com/1996/02/25/business/your–taxes–some–new–tricks–to–help–filers–avoid–an–old–audit–trap.html?pagewanted=print&src=pm. Accessed on 17 May 2011.

Johnston, D. C. (2006). I.R.S. Is Sued on Failure to Release Tax Data. *New York Times*, January 10. http://www.nytimes.com/2006/01/10/politics/10irs.html?sq=IRS%20is%20sued%20on%20failure%20to%20release%20tax%20data. Accessed on 17 May 2011.

Johnston, D. C. (2008). I.R.S. Said to Flout Orders to Yield Data About Audits. *New York Times*, February 12. http://www.nytimes.com/2008/02/12/business/12tax.html. Accessed on 17 May 2011.

Jones, L. (2000-2001). Strange words - audit and examination. *Journal of Tax Practice & Procedure 2*, December–January, 5–6.

Kahneman, D. and A. Tversky (1979). Prospect theory: An analysis of decision under risk. *Econometrica 47*(2), 263–91.

Kim, C. K., J. H. Evans, and D. V. Moser (2005). Economic and equity effects on tax reporting decisions. *Accounting, Organizations and Society 30*(7-8), 609–625.

Kim, C. K. and W. S. Waller (2005). A behavioral accounting study of strategic interaction in a tax compliance game. *Experimental Business Research*, 35–56.

King, S. and S. Sheffrin (2002). Tax evasion and equity theory: An investigative approach. *International Tax and Public Finance 9*(4), 505–521.

Kleven, H. J., M. Knudsen, C. Kreiner, S. Pedersen, and E. Saez (2011). Unwilling or unable to cheat? evidence from a tax audit experiment in denmark. *Econometrica 79*(3), 651–692.

Landsman, W. R., B. L. Miller, K. Peasnell, and S. Yeh (2011). Do investors understand really dirty surplus? *The Accounting Review 86*(1), 237–258.

Leviner, S. (2008). An overview: A new era of tax enforcement - from "big stick" to responsive regulation. *Regulation & Governance 2*, 360–380.

Loomes, G. (2005). Modelling the stochastic component of behaviour in experiments: Some issues for the interpretation of data. *Experimental Economics 8*(4), 301.

Louis, H. and A. Sun (2010). Investor inattention and the market reaction to merger announcements. *Management Science 56*(10), 1781–1793.

McFadden, D. (1973). Conditional logit analysis of qualitative choice behavior. In *Frontiers of Econometrics*. New York: Academic press.

McFadden, D. (2001). Economic choices. *American Economic Review 91*(3), 351–378.

McKelvey, R. D. and T. R. Palfrey (1995). Quantal response equilibria for normal form games. *Games and Economic Behavior 10*(1), 6–38.

Mills, L. F., L. A. Robinson, and R. C. Sansing (2010). FIN 48 and tax compliance. *Accounting Review 85*(5), 1721–1742.

Mills, L. F. and R. C. Sansing (2000). Strategic tax and financial reporting decisions: Theory and evidence. *Contemporary Accounting Research 17*, 85–106.

Moser, D. V., J. H. Evans, and C. K. Kim (1995). The effects of horizontal and exchange inequity on tax reporting decisions. *Accounting Review 70*(4), 619–634.

Obama, B. (2009). Transparency and Open Government - Memorandum for the Heads of Executive Departments and Agencies. January 21. http://www.whitehouse.gov/the_press_office/TransparencyandOpenGovernment. Accessed on 17 May 2011.

OECD (2004). Compliance Risk Management: Audit Case Selection Systems. Prepared by Forum on Tax Administration Compliance Sub-group. http://www.oecd.org/dataoecd/44/36/33818568.pdf. Accessed on 22 May 2011.

Ornstein, C. and H. Limor (2011). Where's the transparency that Obama promised? *The Washington Post*, March 31. http://www.washingtonpost.com/opinions/wheres–the–transparency–that–obama–promised/2011/03/31/AFipwHCC_print.html. Accessed on 17 May 2011.

Pentland, B. T. and P. Carlile (1996). Audit the taxpayer, not the return: Tax auditing as an expression game. *Accounting, Organizations and Society 21*(2/3), 269–287.

Phillips, M. D. (2010). Taxpayer response to targeted audits. Working Paper (University of Chicago).

Rieskamp, J. (2008). The probabilistic nature of preferential choice. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 1446–1465.

Rosenbaum, P. R. (2002). *Observational Studies, 2nd ed.* New York: Springer-Verlag.

Saxon, W. (1994). Philip H. Long, 78; Forced Tax Agency To Be More Open. *New York Times*, July 12. http://www.nytimes.com/1994/07/12/obituaries/philip–h–long–78–forced–tax–agency–to–be–more–open.html. Accessed on 17 May 2011.

Schipper, K. (2007). Required disclosures in financial reports. *Accounting Review 82*(2), 301–326.

Shackelford, D. A. and T. Shevlin (2001). Empirical tax research in accounting. *Journal of Accounting and Economics 31*(1-3), 321–387.

Slemrod, J. (2007). Cheating ourselves: The economics of tax evasion. *Journal of Economic Perspectives 21*(1), 25–43.

Slemrod, J., M. Blumenthal, and C. Christian (2001). Taxpayer response to an increased probability of audit: evidence from a controlled experiment in minnesota. *Journal of Public Economics 79*(3), 455–483.

Slemrod, J. and S. Yitzhaki (2002). Tax avoidance, evasion, and administration. In A. J. Auerbach and M. Feldstein (Eds.), *Handbook of Public Economics*, Volume 3, pp. 1423–1470. Elsevier.

Tan, F. and A. Yim (2011). Appendix to 'Can Transparency Hurt? An Experiment on Whether Disclosure of Audit Policy Details Reduces Tax Compliance.' Unpublished.

The Associated Press (2011). PROMISES, PROMISES: Little transparency progress. *The Washington Post*, March 14. http://www.washingtonpost.com/wp–dyn/content/article/2011/03/14/AR2011031400630_pf.html. Accessed on 17 May 2011.

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

United States Courts of Appeals for the Ninth Circuit (2010). Memorandum for: SUSAN LONG V. IRS, No. 08-35672. September 16. http://www.ca9.uscourts.gov/datastore/memoranda/2010/09/16/08-35672.pdf. Accessed on 18 May 2011.

Wilcox, N. T. (2011). Stochastically more risk averse: A contextual theory of stochastic discrete choice under risk. *Journal of Econometrics 162*(1), 89–104.

Yaniv, G. (1999). Tax compliance and advance tax payments: A prospect theory analysis. *National Tax Journal 52*(4), 753–764.

Yim, A. (2009). Efficient committed budget for implementing target audit probability for many inspectees. *Management Science 55*(12), 2000–2018.

Yitzhaki, S. (1974). A note on 'Income tax evasion: A theoretical analysis'. *Journal of Public Economics 3*(2), 201–202.

Zimbelman, M. F. and W. S. Waller (1999). An experimental investigation of auditor-auditee interaction under ambiguity. *Journal of Accounting Research 37*, 135–155.

Figure 1: Dynamics of underreporting rate over 30 periods

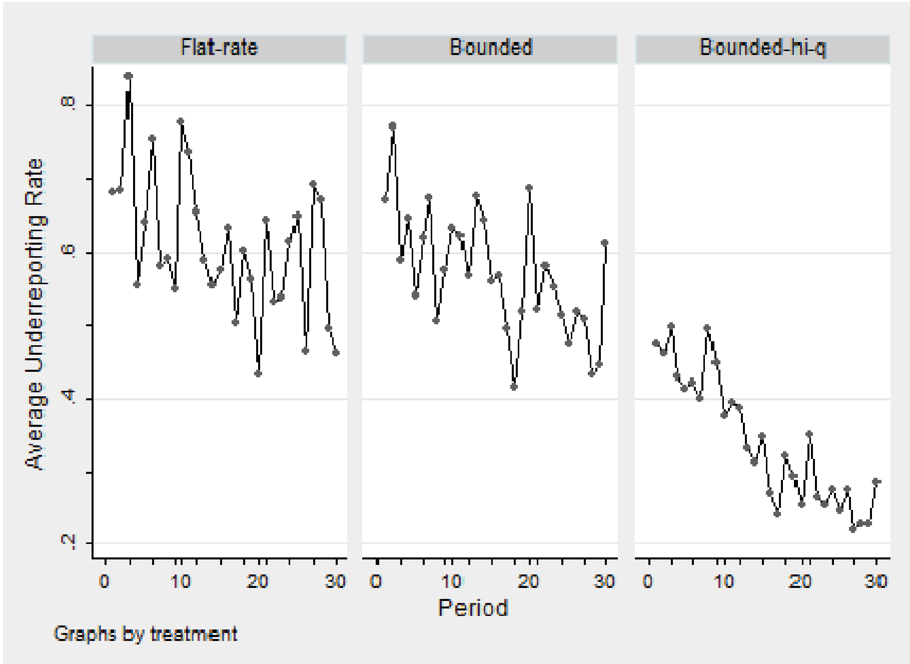

Figure 2: Frequency distributions of individual underreporting rate

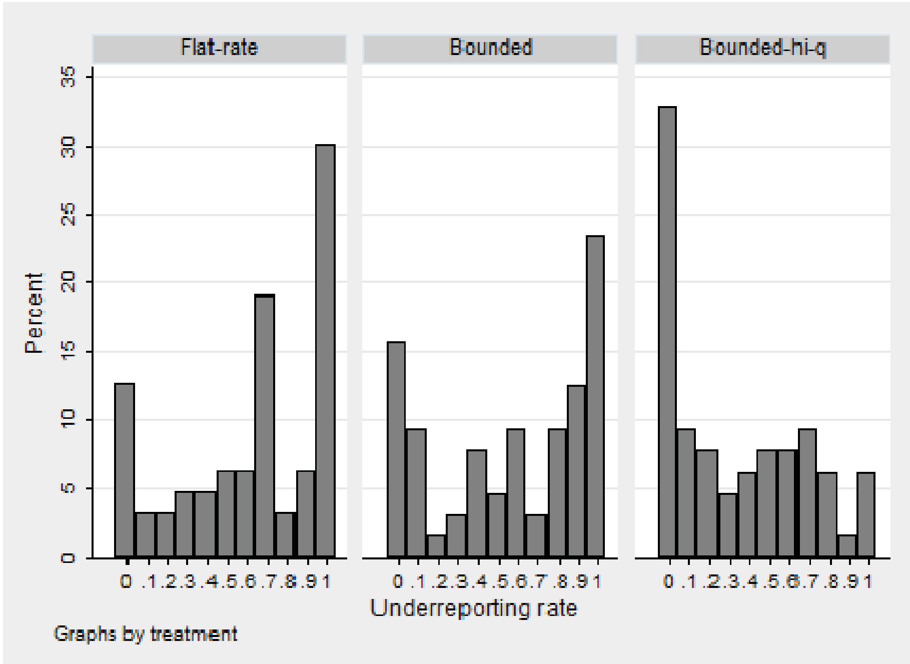

Figure 3: Model predictions of underreporting rate versus actual observations

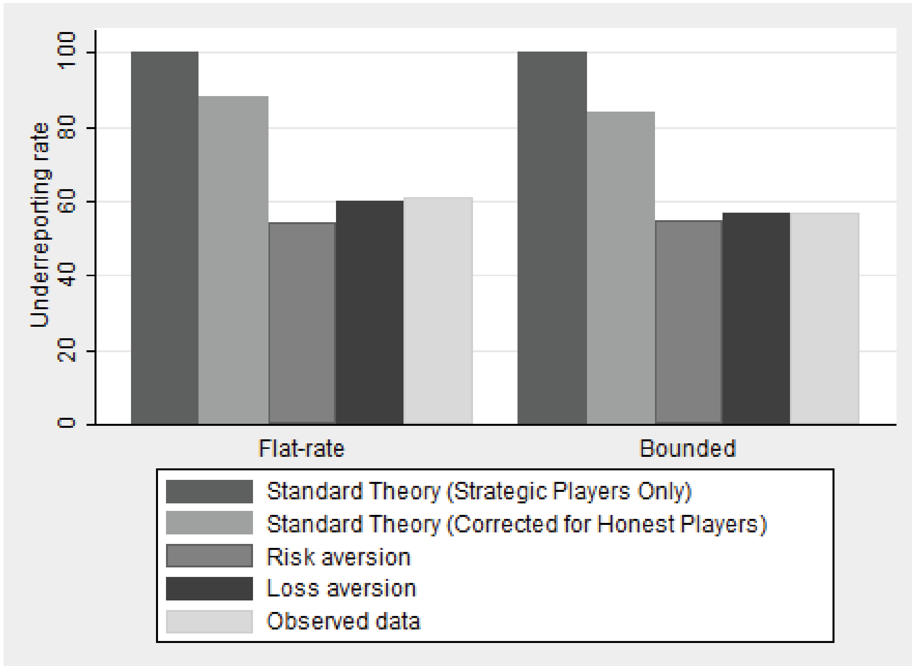

Table 1: Experimental treatments

Treatment	High-income probability q	Audit probability a or capacity K	Number of subjects
Flat-rate	0.5	$a = 0.4$	64
Bounded	0.5	$K = 2$	64
Bounded-hi-q	0.9	$K = 2$	64

Table 2: Summary statistics of treatments

	All 30 Periods			Last 10 Periods		
	Flat-rate	*Bounded*	*Bounded*-hi-q	*Flat-rate*	*Bounded*	*Bounded*-hi-q
All subjects						
High-income frequency	0.514	0.491	0.898	0.527	0.519	0.908
	(0.007)	(0.039)	(0.024)	(0.042)	(0.038)	(0.013)
Percentage of "low-income" reports	79.74%	78.85%	40.31%	77.97%	75.94%	32.97%
	(0.074)	(0.015)	(0.053)	(0.066)	(0.018)	(0.055)
H-type subjects						
Underreporting rate	60.83%	57.11%	33.95%	58.16%	53.32%	26.16%
	(0.144)	(0.049)	(0.038)	(0.143)	(0.052)	(0.046)
Bounded v. Flat-rate	$p = 0.386$			$p = 0.564$		
Bounded-hi-q *v. Bounded*		$p < 0.05$			$p < 0.05$	
Auditing statistics						
Cheater detection rate	38.76%	33.13%	73.27%	42.08%	31.88%	70.97%
	(0.065)	(0.043)	(0.025)	(0.107)	(0.125)	(0.105)
Bounded v. Flat-rate	$p = 0.113$			$p = 0.149$		
Bounded-hi-q *v. Bounded*		$p < 0.05$			$p < 0.05$	
Total no. of audits	153.8	120	114.8	53.75	40	37
	(18.14)	(0.000)	(4.500)	(8.098)	(0.000)	(2.160)
Bounded v. Flat-rate	$p < 0.05$			$p < 0.05$		
Bounded-hi-q *v. Bounded*		$p < 0.05$			$p < 0.05$	
Avg. no. of audits (per group per period)	2.56	2	1.91	2.69	2	1.85
	(0.300)	(0.000)	(0.065)	(0.414)	(0.000)	(0.093)
Bounded v. Flat-rate	$p < 0.05$			$p < 0.05$		
Bounded-hi-q *v. Bounded*		$p < 0.05$			$p < 0.05$	
Audit selection rate	40.16%	31.71%	59.99%	42.96%	32.94%	71.31%
	(0.030)	(0.006)	(0.062)	(0.038)	(0.007)	(0.095)
Bounded v. Flat-rate	$p < 0.05$			$p < 0.05$		
Bounded-hi-q *v. Bounded*		$p < 0.05$			$p < 0.05$	
Budget usage ratio	32.03%	100%	95.63%	32.09%	100%	92.54%
	(0.181)	(0.000)	(0.032)	(0.181)	(0.000)	(0.033)
Bounded v. Flat-rate	$p < 0.05$			$p < 0.05$		
Bounded-hi-q *v. Bounded*		$p < 0.05$			$p < 0.05$	

Note: Standard errors are in the parentheses. Statistical testing on the treatment effects is based on the two-sided Wilcoxon rank-sum test (also called the Mann-Whitney test), with each session constituting an independent observation. *High-income frequency* is the actual frequency of the subjects being assigned as a high-income taxpayer in a treatment. *Percentage of "low-income" reports* is the total number of "low-income" reports received divided by 8, regardless of whether the reports are submitted by genuine low-income taxpayers or lying high-income taxpayers. *Underreporting rate* is the percentage of times where subjects when assigned as a high-income taxpayer submit a "low-income" report. *Cheater detection rate* is the frequency that a tax cheater is caught during an audit stage. *Audit selection rate* is the proportion of "low-income" reports selected for audit, out of the total number of such reports received. *Budget usage ratio* is the percentage of audit resources actually used, out of the budget commitment required to credibly support an audit rule.

Table 3: Estimation of choice models under uncertainty

	Risk aversion		**Loss aversion**		**Loss aversion with Prob. Weighting**	
	Flat-rate	*Bounded*	*Flat-rate*	*Bounded*	*Flat-rate*	*Bounded*
CRRA coefficient r	0.366 (0.350)	0.594 (0.055)				
Gain domain curvature α			0.445 (0.034)	0.428 (0.038)	0.640 (0.459)	0.533 (0.075)
Loss domain curvature β			0.548 (0.052)	0.708 (0.030)	0.586 (0.068)	0.858 (0.073)
Loss aversion coefficient λ			1.100 (0.802)	1.148 (0.030)	1.674 (0.123)	1.283 (0.171)
Weighting parameter δ					1.150 (0.193)	0.899 (0.120)
Perceived audit prob. $\widehat{a}$		0.336 (0.017)		0.305 (0.007)		0.240 (0.023)
Log-likelihood	-1163.773	-1087.292	-1141.710	-1082.473	-1141.353	-1082.111
Observations	2331	2287	2331	2287	2331	2287

Note: All coefficient estimates in this table are statistically significant at the 1% level. To account for within-group correlation, the standard errors are clustered by individual. The *risk aversion* specification is based on a constant relative risk aversion (CRRA) utility function: $u(\pi) = (\pi^{1-r})/(1-r)$, where π is the disposable income (i.e., after-tax income) and r is the CRRA coefficient. The *loss aversion* specification is based on Tversky and Kahneman (1992)'s specification of the value function: $v(x) = x^{\alpha}$ if $x \geq 0$, and $v(x) = -\lambda(-x)^{\beta}$ if $x < 0$, where α and β are the parameters determining the curvature of the function in the gain and loss domains, respectively, and λ is the coefficient of loss aversion. The *loss aversion with probability weighting* specification is based on a popular one-parameter probability-weighting function: $w(\gamma) = \gamma^{\delta}/(\gamma^{\delta}+(1-\gamma)^{\delta})$, where γ is a probability and $\delta \geq 0$ is the weighting parameter. The *perceived* audit probability $\widehat{a}$ is the audit probability of a flat-rate rule that would induce the same level of compliance as observed in the *Bounded* treatment.