An Experimental Study of Taxpayer Compliance Behavior Under Alternative Reporting Regimes

M.S. Plan B Paper

University of Minnesota Applied Economics Graduate Program

Matthew Bombyk

Advisor: Dr. Laura Kalambokidis

1 Introduction

Underreporting of income is a costly problem for the government and for those people who do pay their taxes, due to the necessity of higher tax burdens to sustain a given amount of revenue. An extensive research report published by the IRS this year estimates that in the United States in 2006 the "tax gap" between paid taxes and legally owed taxes was $450 billion, which means 16.9 percent of total tax liabilities were evaded that year (Black et al., 2012). The IRS recovered $65 billion from late payments and audits, but that still left 14.5 percent noncompliance. Breaking the tax gap down into finer categories, the IRS finds that the vast majority (84 percent) comes from underreporting of income, most of that (62.5 percent) comes from individual income taxes, and most of that (52 percent) is small business proprietor's income. This totals to 27 percent of noncompliance due to underreporting of individual proprietor income. It is estimated that 57 percent of business income is not reported (Black et al., 2012). Wages make up a small fraction of underreporting, mostly due to the fact that firms must report employee income directly to the IRS, and withholding is common, so a relatively disinterested third party makes the decision of how much income is reported (Slemrod, 2008). Slemrod emphasizes the importance of enforcement in compliance behavior, citing the fact that income subject to withholding and substantial information reporting (wages) has a 1 percent noncompliance rate, compared to 56 percent noncompliance for income with little or no information reporting requirements.

Given the difficulty of identifying cheaters and the cost of increasing the rate of auditing, a better understanding of the determinants of noncompliance is needed to reduce the magnitude of tax cheating. To address a part of the tax evasion quandary, Kalambokidis et al. (2012) conducted a laboratory experiment[1] which was designed to examine the feasibility of using the choice between a high-burden,[2] low-transparency and a low-burden, high-transparency tax regime, as a mechanism to separate out those who have higher and lower propensities to cheat when reporting their income. The results are the subject of this paper.

Examples of existing mechanisms with a similar structure to this experiment in terms of burden-transparency tradeoff include the Compliance Assurance Process (CAP) in the United States. This is a program that large businesses can choose to enroll in, and in which the business works cooperatively with the IRS to identify potential issues in the tax reporting process, and resolve them before filing. In return, the businesses face less invasive and extensive audits post-filing. Businesses in this program have higher income transparency since the IRS examines their operations closely, but faces a lower reporting burden in the sense of less costly audits and less chance of needing to re-file. Other examples include the Gift Aid program in the U.K., the Tip Rate Determination Agreement (TRDA), Tip Reporting Alternative Commitment (TRAC), the Employer Designated Tip Reporting Alternative Commitment (TRAC), and the recently discontinued Attributed Tip Income Program (ATIP), all from the IRS.

The remainder of this paper is structured as follows: Section 2 reviews the related

[1] I was a research assistant for this experiment.

[2] Burden in this context is "reporting compliance burden," which is the transaction cost of reporting income to the tax authority, and therefore includes the costs of time and stress expended filling out paperwork, or the services of a tax preparer. Transparency refers to the extent to which earned income can be hidden from the tax authority within a particular tax-reporting regime.

literature on tax compliance and use of regime choice as a separating mechanism. Section 3 explains the research questions addressed in this paper. Section 4 describes the experiment and discusses the methodology used in the analysis. Section 5 presents results from the study, and Section 6 concludes.

2 Literature Review

The economic analysis of tax compliance decisions by rational taxpayers was formally initiated by Allingham and Sandmo (1972), based on the economics of crime model first expounded by Gary Becker (1968). In the Allingham/Sandmo model, taxpayers begin with an exogenous amount of income, and then choose how much to report to the tax authority to maximize expected utility, subject to an exogenously determined audit rate, penalty rate, and tax rate. Formally the problem can be stated as

$$\max \mathrm{E}\left[U(X)\right] = (1-\alpha)U(Y - tX) + \alpha U((1-t)Y - \pi t(Y - X))$$

$$\text{subject to } 0 \leq X \leq Y$$

In this expression, X is the amount of reported income, Y is actual earned income, t is the tax rate, α is the audit rate, and π is the penalty rate. Comparative static analysis reveals that increases in both audit and penalty rates reduce underreporting, holding other variables constant. An exogenous increase in income will increase the fraction of income underreported, given decreasing risk aversion,[3] but will decrease or

[3]Specifically the Arrow/Pratt coefficient of absolute risk aversion, $R_A(y) = -\frac{U''(y)}{U'(y)}$

stay constant with increasing or constant risk aversion, respectively. If we make the assumption that risk aversion is decreasing in income,[4] then one implication of this model is that effective tax rates are made more regressive because of noncompliance.

The effect of an official tax rate increase on the amount of income underreported is ambiguous and may actually decrease underreporting if absolute risk aversion is everywhere decreasing in final[5] income y. This tax rate effect was surprising when first derived because as tax rates rise, the marginal benefit of underreporting increases. However, the full tax rate effect can be decomposed into two parts:

$$\frac{\partial X^*}{\partial t} = \text{Substitution Effect} + \text{Income Effect} \tag{1}$$

The first term is a "substitution effect," which is always negative and reflects the direct marginal effect of increased tax rate: higher tax rate means you gain more from each dollar not reported. The second term is the "income effect" which is positive, zero, or negative as R_A is decreasing, constant, or increasing. Essentially, with risk averse preferences the income gained from additional underreporting is weighted less in the utility function than the income lost on that underreporting if they are audited. The overall effect of this depends on audit probabilities, the penalty rate, the tax rate, and the curvature of the utility function. Using decreasing R_A as our assumption, the income effect is positive. This makes the sign of the tax rate effect ambiguous.

A simple extension of this model to a more realistic tax structure was carried

[4]This is a standard assumption in the economics of uncertainty (Sandmo, 2012)

[5]"Final income" is post–tax-and-penalty income

out by Yitzhaki (1974), who modified the penalty structure considered. Allingham & Sandmo assumed the penalty was linear in the amount underreported. However, it is more commonly the amount of *evaded tax* that is penalized, for example in the United States. When penalties are made linear in the amount of tax evaded, the effect of tax rate is no longer ambiguous given decreasing risk aversion. In particular, the substitution effect in Equation (1) disappears, so higher tax rates increase underreporting due to the income effect. The effect of a tax rate increase on total revenue is also ambiguous in this situation. Since we use the penalty structure studied by Yitzhaki in this experiment, henceforth I will refer to this model as the Allingham-Sandmo-Yitzhaki (ASY) model.

The most pronounced empirical regularity with respect to this model is the fact that the ASY model predicts a dramatically higher level of tax avoidance than is actually observed in most countries. As mentioned above, the most extensive studies of tax avoidance have been carried out by the IRS, which estimates that about 91 percent of taxable income is reported annually to the IRS (Andreoni et al., 1998). In contrast, using approximations of the current tax, audit, and penalty rates, and a commonly-used functional form for utility, the ASY model predicts not only that all taxpayers will underreport at least a small amount, but that the level of relative risk aversion needed to attain current levels of income reporting in the United States is close to 30, whereas in reality this parameter is typically estimated to lie between 1 and 2 for most individuals (Alm, 2012).

Several studies have investigated how individual differences among taxpayers can be used by a tax authority to target their audits more effectively. Falkinger & Walther

(1991) propose offering taxpayers a choice between two tax regimes: the first is the prevailing tax regime (NO BONUS), and the alternative (BONUS) offers a lower tax rate but with a higher penalty for underreporting income. They use the ASY model as their baseline, and thus all taxpayers are supposed to have some positive propensity to cheat, moderated only by their level of risk aversion. Taxpayers are distinguished by the amount that they choose to underreport under the prevailing tax regime. An arbitrary cutoff is proposed, where those who underreport more than the cutoff are considered "big fish" and those who underreport less "small fish." Falkinger & Walther show that if the government raises audit rates in the NO BONUS regime, the big fish who remain are induced to pay higher taxes, while those who switch to the BONUS regime also pay more taxes due to the deterrence effect of the higher penalty rate. Though small fish may pay less tax, overall revenues to the tax authority increase if the audit rate is set sufficiently high in the NO BONUS regime.

Raskolnikov (2009) also proposes two regimes that taxpayers must choose between, after which the government can target enforcement efforts. Raskolnikov explicitly acknowledges the various economic and non-economic reasons that different individuals have for paying (or not paying) their taxes, such as fear of disapproval, guilt, duty, and habit, and proposes a tax structure that aims to separate those with different motivations for paying their taxes. See Kalambokidis et al. (2012) for a more extensive summary of this complex proposal.

The first regime (deterrence regime) in Raskolnikov (2009) is similar to the existing tax system, but with steeper penalties for underreporting. The high penalties serve to dissuade rational ASY-types ("gamers") from underreporting too much, since high penalty rates reduce noncompliance for this group. The second regime (compliance regime) has low penalties for underreporting but involves a set of "conditions." The conditions should be very costly for gamers, but relatively painless for "non-gamers," those with other motivations for paying taxes. One example given is the requirement to agree that in any future disputes with the tax authority, the taxpayer will essentially be treated as "guilty until proven innocent" and the burden falls on the taxpayer to provide clear evidence supporting their position. The compliance regime would also include enticements such as "readily available prefiling advice, respectful audits, mass media campaigns emphasizing widespread compliance" and similar measures that make life easier for those who do not wish to cheat, but provide relatively minor benefits for gamers. A key feature of this model is the use of legal and other means that are not traditionally "economic" to create an effective sorting mechanism.

Though Raskolnikov (2009) and Falkinger & Walther (1991) propose regime choice mechanisms that could theoretically improve overall compliance by separating taxpayers with different characteristics, neither tests whether their proposed mechanisms actually work to separate different types of taxpayers. This study is a first step in filling that gap. While the complex legal components of the Raskolnikov proposal are difficult to implement in a laboratory, the economic parameters studied in the ASY model and by Falkinger & Walther are simple to manipulate. Consequently, those forces are what the design of this study was based on. However, in this study, rather than providing the option of a low-tax, high-penalty regime to taxpayers, we provide a low-burden, high-transparency regime. The low-burden aspect can be interpreted similarly to Raskolnikov's proposal to offer tax filing services to taxpayers in the compliance regime, while the high-transparency aspect means that underreporting is very difficult, and in this experiment it is actually impossible.

3 Research questions

This paper addresses a number of research questions. The first questions can be answered within the relatively narrow context of our experimental setting. First, is subjects' behavior consistent with standard economic theory, and if not, how does it deviate? Second, does propensity-to-cheat (PTC), as measured using tax reporting behavior in an earlier stage of our experiment, affect the choice of tax regime in a later stage?

The next set of questions concerns the experimental setting itself, and addresses the study's external validity. First, are subjects' cheating behaviors in the lab consistent with their self-reported tax morale,[6] as measured by the World Values Survey?[7] The idea behind this question is to determine how much of what we are measuring as PTC is due to some underlying "true" propensity to cheat on actual taxes, and how much is restricted to the context of our lab experiment, and thus may be measuring something closer to general risk aversion, since the experimental setting lacks some of the features of most real-world taxes such as contributing to a public good. Next, I want to investigate more closely the subset of our subjects who are business own-

[6]Tax morale is defined as "an individual's innate willingness to pay taxes."

[7]See http://www.wvsevsdb.com/wvs/WVSAnalizeQuestion.jsp

ers or partners. These individuals are empirically much more likely to underreport income than the general taxpayer population, and our experiment may be able to shed light on whether business owners are "naturally" more prone to cheating, or whether it is the differing opportunities to cheat that cause the large differences in tax compliance between these groups. Though our sample of business owners is small (34 subjects), we can provide preliminary evidence on this topic.

4 Experimental Design & Methodology

4.a Experimental Design

There are several factors that cause people to differ in the amount that they underreport income, including different opportunities to cheat, different propensities to cheat, the enforcement environment (e.g., audit and penalty rates), and income level. The experimental approach allows us to hold the opportunity to cheat constant across individuals, vary the enforcement environment, constrain income variation to a small band, and measure propensity to cheat by observing subjects' underreporting choices in our experimental setting. The measurement of cheating propensity, in particular, is essentially impossible in a natural setting, because even if an extensive and randomly assigned auditing campaign could accurately identify the actual cheating rates of different types of people, other correlated factors (omitted variables) will confound any attempt to attribute these different cheating rates to underlying differences in "propensity to cheat." Lab experiments are often subjected to criticism for failing to possess external validity, partly due to the overrepresentation of university students in experimental subject pools. Due to the heavy concentration of tax evasion among business owners, we attempted to recruit from a non-student population as much as possible, and 10 percent of subjects were business owners or partners, while 28 percent were non-students.

The experiment consisted of three stages, and each stage consisted of a number of rounds. In each round, subjects performed a simple word sorting task on a computer to earn "income," and then reported that income to the experimental tax authority.[8] The task involved sorting a list of 30 randomly selected words into alphabetical bins, and the number of words the subject correctly sorted within the time span of the round was multiplied by a "wage rate" to determine income. The amount of time to earn income and the wage rate were held constant across rounds, while the other parameters were varied, as is detailed below.[9] Table 1 gives a summary of the different experimental "states" faced by each subject. Each state is characterized by a particular combination of parameters. Prior to each stage, subjects read instructions and answered three quiz questions to test their comprehension,[10] and also completed one practice round.

In the first stage, subjects faced 12 different combinations of tax rates, audit rates, and penalties for underreporting income in each round. All subjects saw the same combinations of parameters, but in random order. After completing the sorting task, subjects chose how much income to report. A screen was displayed where subjects could enter different amounts of reported income, and before submitting their report

[8]The "tax authority" was actually just the computer that a subject was using, which internally recorded the amount reported.

[9]Earning time was 30 seconds per round and wage rate was $1.20 per task completed.

[10]Subjects were paid $.50 for each correctly answered quiz question, for a total of up to $4.50.

they were able to see post–tax-and-penalty income in the case of both audit or no audit. Tax was removed from the reported income, and if a subject was audited and they had underreported, they paid the full tax plus a penalty, which was calculated as the penalty rate multiplied by the amount of evaded tax, $\pi \times t(Y - X)$. From this stage we were able to construct measures of each subject's propensity to cheat, which are detailed below.

In the second stage, subjects reported their income from the earning task by filling out a reporting form that involved locating a number on their carrel,[11] entering two numbers that were presented on the screen (sorting tasks completed and seconds per task), and multiplying the wage rate by the number of tasks completed to calculate total earnings. Before beginning the round, however, the subjects could choose to pay a "burden reduction fee," which allowed them to skip the reporting form. There was no opportunity for underreporting whether they paid the fee or not. In each of the five rounds subjects faced a different burden reduction fee, ranging from \$0.25 to \$4.00. Stage 2 allowed us to construct indices of willingness-to-pay for burden reduction, which we use as controls for the subjective cost of reporting (reporting burden) in later analyses.

In the third stage, prior to the earning task in each round, subjects had to make a choice between two tax reporting regimes. In the "automatic reporting" regime, subjects had to fill out a form, but after that form was filled out they were done entering information and the computer recorded their actual income, leaving no opportunity to cheat. The form type in the automatic regime could take the values

[11] A carrel consists of a desk with a computer screen on top, along with separators on both sides to shield each subjects' screen from other subjects. It looks like a cubicle or a voting booth.

of NONE, SHORT, or LONG. NONE involved no form to fill out, SHORT involved only entering the carrel number, and LONG is as described for Stage 2. In the "self reporting" regime, subjects were always required to fill out the LONG form, but then reported their own income as in Stage 1, which gave them a chance to underreport income. The third stage allows us to examine subjects' behavior when faced with tradeoffs between tax reporting burden reduction and the ability to underreport income, and also the effectiveness of using the choice of tax reporting regime as a proxy for propensity to cheat.

At the end of the experiment, one round from each stage was randomly selected to be the payout round. The subject was then paid half of the post-tax and penalty dollar amount earned in each of the three selected rounds, plus the payment for correctly answered quiz questions.

4.b Theoretical framework

The Allingham/Sandmo/Yitzhaki model is the basic theoretical framework for making and testing predictions about our subjects' behavior. As detailed above, this is a standard expected utility model applied to the case of tax evasion where the penalty is linear in the amount of tax evaded. Here I apply the model to the specific parameter values in our experiment to obtain predictions about observable behavior.

I will first consider the simplest case of a subject with risk-neutral preferences, i.e., a utility function that is linear in income. This person will attempt to maximize the expected value of their payoffs. Let α be the audit rate, π the penalty rate, t the tax rate, Y actual earnings, X reported earnings. Then the taxpayer's problem is:

$$\max_{0 \leq X \leq Y} \text{EV}(X) = (1-\alpha)(Y - tX) + \alpha((1-t)Y - \pi t(Y - X)) \tag{2}$$

Consider the derivative of the objective function with respect to amount reported, $\frac{d\text{EV}(X)}{dX} = \alpha\pi t - (1-\alpha)t$. Note that this is a constant. If this expression is < 0 then an additional dollar reported always reduces the expected payoff. The subject has the incentive to underreport the maximum amount, so $X^* = 0$, where X^* is the maximizer. If $\alpha\pi t - (1-\alpha)t > 0$ then an additional dollar reported always increases the expected payoff. The subject will fully report, so $X^* = Y$. Only when the probability-weighted marginal benefit of underreporting exactly equals its probability-weighted marginal cost, so that $\alpha\pi = 1 - \alpha$, will an interior solution exist. In this case any report is optimal.

Plugging our experimental parameter values into this formula, a risk-neutral subject will choose either to report no income (states 1-8, 18-23), report any amount of income (states 9-10), or report all income (states 11-12, and 24-26). Note that this matches up closely with the audit rate. For audit rates of 0 and 0.1 it is always optimal to fully underreport. For audit rate of .5, any amount is optimal when the penalty rate is 1, and full reporting is optimal when the penalty rate is 2.

As risk-aversion increases, rational agents will put less of their money into a "risky asset." In this case the risk-free asset is the amount of income reported, while the amount underreported is the risky asset. So strictly risk-averse subjects will underreport less (report more of their income) than risk-neutral subjects. A risk-averse subject will still report all income in the negative-expected-value states. They will also report all income for the two previously ambiguous states (in these cases underreporting is a "fair gamble" which is always rejected by a strictly risk-averse agent). For the zero-audit-rate rounds they will still report zero income, since these rounds do not truly represent a choice under uncertainty: any income not reported is just free money. They may or may not report zero income in the remaining states however (those with audit rate=.1); the prediction is ambiguous and depends on the level of risk aversion of each individual subject. However, since the expected value of underreporting is always positive in these rounds, no rational agent will fully report their income (Varian, 1992, p. 184).

4.c Construction of propensity-to-cheat and willingness-to-pay indices

We constructed two indices to capture the extent of subjects' propensity to cheat. The first index, PTC1, is the fraction of rounds (out of 12) within Stage 1 that an individual reported less than their full income. We might say this measures the breadth of cheating. The second, PTC2, is the fraction of total income earned in Stage 1 that was not reported. This index measures the depth of cheating. The two indices correlate at 0.77, so there is substantial independent variation between the two. We compare our results below using either index, to see if they are sensitive to the particular measure of PTC used. Since PTC is an ill-defined latent variable, we have no strong priors about which index is "better" than the other, and in future work constructing a composite of the two may be useful. PTC is a hybrid concept encompassing both non-economic forces such as "tax morale" and economic forces like risk aversion.

Our measure of willingness-to-pay for burden reduction (WTP) is more straightforward than the PTC indices. The WTP index is an attempt to proxy for subjective cost of reporting, but WTP is an objective quantity that we measure in dollars. We do not have a perfect measure since people are not forced to reveal their true WTP but rather we see which of several dollar amounts are below and above their actual value. We elected to use a WTP index that is the fraction each subject paid, out of the total amount that they could have paid ($7.75).

4.d Regression analysis

Predicting underreporting in Stage 1

To investigate more carefully the effect of the experimental parameters on underreporting of income, I employed two forms of Tobit regression, namely a two-limit Tobit and an extension of this to a random-effects panel model. The dependent variable in these models was the fraction of income underreported. The Tobit model is useful when you face a substantial number of subjects facing a corner solution. In the case of underreported income, there were actually two corner solutions: subjects could not underreport less than zero or more than their total income.[12] Since the level of income varies by subject and round, converting the amount of underreporting to a fraction of total income simplifies the analysis considerably.

In a situation with corner solutions, a standard ordinary least squares estimate

[12] Technically negative underreporting was possible. However, it provided no benefit and was rare (3% of all reports). For this paper, negative amounts were revised to equal zero prior to analysis.

of the slope will generally be biased downward (Wooldridge, 2010, pp. 674-5). Tobit surmounts this problem by assuming the existence of a latent index variable that is "censored" in the case of a corner solution. In our case when this latent variable takes a value below zero or above 1, then we observe 0 or 1, respectively.

The model further assumes that the distribution of errors in this model is normal, and its estimates are inconsistent if this assumption fails to hold. However, there is no obviously better alternative for estimating the conditional mean in this context that does not make some questionable assumptions, and since our experimental setting is quite different from any real-world context, it is more important to get a general idea of the direction and relative magnitudes of the effects of experimental parameters, rather than trying to precisely estimate some empirically important quantity. I decided that controlling for the corner solutions in an imperfect way was better than just ignoring them. One thing to note about the regressions is that the estimated effects of the experimental parameters should be invariant (within random error) to which control variables are included in the model, since they are entirely independent by virtue of the experimental design. Thus control variables are included only to improve estimates of the effect of income, business-owner status, and the cheating question from the World Values Survey.

Predicting regime choice using PTC and WTP

To investigate the effect of propensity-to-cheat and cost of reporting burden on the choice of subjects between the automatic reporting and self reporting regimes, I used the probit regression model. Probit is one of the most commonly used methods for studying the generation of binary dependent variables, and models choice using a linear index of the explanatory variables. When this index plus a normally-distributed random error exceeds zero, the subjects chooses the self-reporting regime and we observe a value of 1 for the dependent variable "regime." Otherwise the value is zero. Estimation of parameter values proceeds by maximum likelihood techniques. As in the case of the Stage 1 Tobit regressions, I estimated both standard probit and random-effects probit models, to account for the panel nature of our dataset.

The key predictors of interest were PTC1 and PTC2 from Stage 1, which entered into separate models, and WTP from Stage 2, which was used to control for differences in the cost of burden. The idea behind controlling for WTP is that two main forces are working in the decision between automatic and self reporting: subjects consider both how much reporting burden they can avoid by choosing automatic reporting and how much they could gain by underreporting income in the self-reporting regime. If PTC and WTP are at all correlated, then failing to control for WTP will cause the regression estimates to falsely attribute the "reporting-burden-effect" to PTC.

In the tables below I report marginal effects, also known as average partial effects. Since probit predictions/estimated probabilities of choosing self-reporting are a nonlinear function (the normal cumulative distribution function) applied to a linear index, the derivatives of the predicted value (i.e., the partial effects) are not constant but depend on the level of the index. As such the coefficients estimated are difficult to interpret. Average partial effects report the average of the partial effect of a variable over all the predicted index values observed in the data.

Finally, I separated the data into subsets based on audit rate (3 rounds for each audit rate), to investigate differences in the decision structure with changes in the expected value of cheating in the self-report round. Initially I separated the regressions by each experimental state, but there were no differences in estimated effects within each audit rate category.

Determining the effect of PTC and WTP on underreporting in Stage 3

I again used Tobit regressions in Stage 3 to predict underreporting contingent on having selected the self-report regime. I compared the results to Stage 1 Tobit estimates using the subsample of Stage 1 with the same parameter values as Stage 3 (penalty rate=2, tax rate=.15). I also used a random-effects Tobit specification, as in Stage 1. These regressions allow one both to check for consistency of behavior between stages (by looking at the coefficient on PTC) and to determine if self-selection into the self-reporting regime changed aggregate patterns of reporting behavior within that regime, compared to the case where there is no choice of regime.

5 Results

5.a Descriptive statistics

Subject demographics. Table 3 gives the demographic breakdown of our subjects. 52 percent of subjects are male. 76 percent are under 30. 72 percent are students, fairly evenly distributed across grades. 36 percent are not employed, while 10 percent are business owners or partners. This table also contains the mean values for the PTC and WTP indices, and the World Values Survey cheating question (WVS index), all of which range between 0 and 1. Generally, both PTC indices and the WVS index decrease with age, while WTP increases. Asians give much higher scores on the WVS index than any other race. Students score higher on both PTC indices and the WVS index, as do individuals who are not employed.

Business Owners. I ran a t-test of three variables, PTC1, PTC2, and the World Values Survey cheating question to determine if business owners have higher propensity-to-cheat than the rest of the population. Three things are notable about the results. First, as can be seen in Table 3, business owners actually have higher mean scores than nonbusiness-owners for the PTC indices, but lower scores on the cheating question. Second, the differences are not significant at any conventional level. Third, using the estimated means and standard deviations as approximations to the actual levels, the power of all three tests is less than 0.2, so the lack of statistical significance gives us little information about the actual relationship. On average, business owners are 6.4 percent more likely to choose self-reporting than non-business-owners based on the Probit specification reported in Table 5. Though this is statistically significant, it is a small difference.

All of this evidence suggests that it is likely the low information reporting requirements for sole proprietors within the current tax system in the United States, rather than fundamental differences in propensity to cheat among business owners, that is the driver of the heavy underreporting of small business income relative to other types of income. Black et al. (2012) discuss the dramatic differences in income underreporting when separating types of income by information reporting requirements.

Another possibility is that business owners are used to making simple calculations and find the self-reporting regime quite easy. The sample size is unfortunately too small to make any conclusions about this group.

5.b Stage 1 Parameter Effects

I take several approaches to analyzing subjects' underreporting behavior in Stage 1 of the experiment. First I attempt to sort subjects into categories based on their behavior. The largest and most obvious category is those who never underreport. These people make up 23 percent of the sample, and I refer to them as "Compliers." It is tempting to further categorize the remainder of subjects based on a more detailed analysis of their Stage 1 behavior, but in this paper I simply classify subjects as Compliers or "Gamers" (everyone that underreports in at least one round). Over all of stage 1, 53 percent of subject-rounds led to zero underreporting, and 17 percent to full underreporting. It is also interesting is to look at how this relationship changes with the audit rate. Figure 1 is a histogram of the fraction underreported in Stage 1, by audit rate. The importance of audit rate in underreporting decisions, as well as the heavy concentration of reporting at either all income or no income, are very apparent in this figure. Figure 2 is a graphical representation of Table 2a, which shows the mean fractions underreported by experimental state in Stage 1. The most notable feature of this graph is the large differences in mean fraction reported by audit rate, compared to much smaller differences by penalty and tax rates. Table 2a reveals that for nonzero audit rates, the penalty rate also makes a difference (4 percent more subjects underreport when the penalty rate is only 1). This is not the case when audit rate is zero which makes sense since penalty does not apply. Tax rate has no discernible effect for any combination of audit and penalty rates. About 36 percent of subjects were fully compliant even when the audit rate is zero. This is higher than suggested by our first PTC index, which indicates that 23 percent of subjects never underreport in Stage 1, while 10 percent always underreport. Perhaps this discrepancy is caused by subjects experimenting with underreports and then switching back to full reporting. The second PTC index is the fraction of earned income underreported over all of Stage 1. The highest level of this index is .84, meaning that apart from the large group of Compliers, everyone else was sensitive to the parameters in the experimentnobody always chose to report zero, even if they did always underreport.

Table 4 presents the regression results from Stage 1. Both the signs and magnitudes of the estimates are interesting. First note that the effect of increasing the audit rate is substantially larger than the effects of varying the other parameters, at least within the ranges we chose to use in the experiment. This is what we saw with the descriptive statistics. Next it is interesting to note the basic conformity with the predictions of the Allingham/Sandmo/Yitzhaki model. As the models predicts, in our experiment increasing both audit rate and penalty rates improves compliance, and higher income decreases compliance. The income effect is substantial considering the standard deviation of this variable was 4.9. Moving from \$15 to \$25 income raised the average fraction of underreported income by 37 percent. The model also predicts a positive effect of increasing the tax rate on compliance if subjects exhibit decreasing absolute risk aversion, and zero effect if a subject is risk-neutral. Although our point estimates of the tax rate effects are positive, they are not significantly different from zero. Based on these regression results our subjects' behavior does not appear to contradict this model. We also note that there is no detectable difference in underreporting between business owners and the rest of the population. The WTP index has no significant effect, suggesting that PTC and WTP are not closely related. This is also corroborated by their small zero-order correlations, which are -.09 and -.13 for PTC1 and PTC2, respectively.

5.c The World Values Survey Cheating Question

We have just seen that in aggregate our subjects behavior appears to fit well with the ASY model. However, this model fails to explain the behavior of Compliers, who never cheat even when the audit rate is zero. Non-economic motivations must be invoked to explain this behavior. This was the inspiration for asking our subjects the World Values Survey question: "Is cheating on your taxes, if you have the chance, always justifiable, never justifiable, or something in between?" Subjects answered this question by indicating a number between 1 and 10. We have rescaled this question to range between 0 and 1, where 0 corresponds to "never justifiable" and 1 corresponds to "always justifiable." This rescaled index will be referred to as the "WVS index."

Tables 5 and 6 and Figures 3 and 4 examine this question further. Table 5a displays the raw correlations between the WVS index and the two PTC indices. Correlations range from .22 to .26 depending on the type of correlation coefficient used. Although these are positive and significantly different from zero, these correlations are quite small and suggest that most of the variation in within-experiment propensity to cheat is coming from sources other than tax morale. There are two obvious explanations for this result. First, we may lose faith in our PTC index and claim that it is not measuring "realistic" tax reporting behavior, which is what the WVS question is specifically framed to capture. Alternatively, we can interpret these results as evidence that the WVS index is not in fact capturing the characteristic that it was designed for, perhaps because it is simply a self-report with nothing on the line. In this case we would be asserting that within-experiment, incentive-driven behavior is more valid at capturing "true" PTC than the self-reported answer to a survey question. Our evidence alone cannot distinguish between these two explanations.

We can use regression to get a clearer picture of the relationship between the WVS index and the PTC indices. Figure 3a graphs the overall effect (linear plus quadratic terms) of the WVS index in the Tobit regressions reported above. This figure clearly shows some nonlinearity in this relationship—at low values of WVS index, increases are associated with substantially more underreporting. This peaks around .55 or .66 and then levels off or falls slightly. The decrease at high WVS index values may just be due to the restrictions imposed by the linear and quadratic functional form. This type of relationship is difficult to interpret given that the question was answered on an inherently ordinal Likert scale, so I will just treat this as an empirical observation.

When we look more specifically at the Complier/Gamer dichotomy, the WVS index comes into much clearer focus. Figure 4 shows the distribution of fraction of income underreported by Complier status. 51% of Compliers answered that cheating was “never justifiable” compared to 21% of Gamers. Table 5b displays the correlations between PTC indices and the WVS index for Gamers only. The correlations drop to between .12 and .15, which are very small correlations, although they are significantly different from zero. This decrease suggests that much of the observed correlation between the PTC and WVS indices comes from “shared zeroes.” In other words, many subjects are both Compliers and think cheating is never justifiable. Figure 3b displays the overall WVS index effect from Tobit regressions using only Gamers. The relationship is almost reversed: now WVS index responses seem to have little effect at the low ranges, but the effect greatly increase as WVS index increases. The large effect at low ranges that we saw in Figure 3a appears to be due almost entirely to Compliers. I also ran a probit regression to predict Complier status using the WVS index and demographic controls. The results are reported in Table 6. Two things are notable about the results. First, they corroborate the story I have been telling about the relationship between the WVS question and Complier status: the marginal effect of the WVS index (essentially, the average effect of moving from 0 to 1) is to decrease the probability of being a Complier by 22%. The second interesting result is that age is the only demographic variable that significantly effects the probability of being a Complier, and being older substantially increases this probability.

Overall these results suggest that there are important links between the WVS question and in-experiment behavior. Moreover, the primary effect is that those who think cheating is never justifiable often never cheat. The WVS question appears to be effectively capturing some of the non-economic factors that cause some individuals to never underreport. Further investigation into this question could give strong support to the external validity of this experiment.

5.d Stage 2 results

The demand for burden reduction is displayed graphically in Figure 5, which is adapted from Kalambokidis et al. (2012). Same information is contained in Table 2b. The demand curve for burden reduction is downward sloping and elasticity increases slightly at higher prices. It is worth noting that 42 percent of subjects pay for burden reduction in at least one round, so clearly the reporting burden is at least somewhat salient for many subjects. Only 10 percent of subjects pay the full \$4.00. I will also note here that Stage 3 average income is \$26. If fully reported, the tax paid on that income at the 15 percent rate is \$3.90, very close to the maximum burden reduction fee. If we consider the case in Stage 3 where the automatic reporting form is NONE, and the audit rate is zero, then the average gain in monetary payout by switching from the automatic reporting regime to the self-reporting regime and reporting zero income is more than the cost of reporting for about 90 percent of subjects. Although we do not specifically look at the consistency of these individuals' behavior across stages, it is an interesting direction for future work with these data.

5.e Predicting regime choice using Stage 1 and 2 behavior

Table 5 presents the average partial effects from several probit specifications predicting reporting regime choice in Stage 3. The first thing to note is the differences in the two models. The random effects model estimates much larger (in magnitude) coefficients than the standard probit, while the signs are the same. Moreover, the significance of the world-vales survey cheating question, and the effect of being a Complier, both disappear. It is not clear to me how to interpret these findings. Further investigation of the correct specification of this relationship, including use of the correlated random effects probit model, may be warranted for future work. I will focus on the standard probit results for the remainder of this section. As anticipated, higher audit rates make self-reporting less likely. Higher income makes self-reporting more likely. Surprisingly, the form type, which is an attempt to vary the amount of reporting burden avoided by switching to the auto-reporting regime, had no effect on subjects' regime choice. On the other hand, willingness-to-pay for burden reduction did substantially increase the likelihood of choosing the auto-reporting regime, as expected. The main coefficients of interest are the propensity-to-cheat indices (PTC1 and PTC2) which both increase the likelihood of choosing self-reporting. This indicates that the choice of regime is causing some self-selection of those with higher propensity-to-cheat into the self-reporting regime. Furthermore, those who never cheated in Stage 1 (Compliers) are approximately 6-9 percent less likely to choose the self-reporting regime, which is the anticipated direction but smaller than expected. We also noted that 35 percent of subjects who chose self-reporting did not actually cheat, and that this result holds across reporting form types. These people present a puzzle, since they chose to accept a higher reporting burden, but neither they nor anybody else benefited in any way from their doing so.

To investigate the choice of regime more closely, I ran an additional set of probit regressions, stratified by audit rate. These are reported in Table 6. The marginal effects of income and WTP do not differ dramatically between audit rates. Form type remains insignificant. Note that all of the significance in the PTC1 coefficient is coming from the round with audit rate = 0.1. This is consistent with an expected-utility interpretation of the results: rational risk-neutral or risk-averse subjects who are attempting to maximize expected utility are going to report fully when audit rate = 0.5, and report nothing when audit rate = 0, but individual differences in risk aversion should lead to reporting differences in the rounds where audit rate = 0.1. Another interesting finding is that "Complier" has a significant effect only in the audit rate = 0 states, where it increases the likelihood of choosing the automatic reporting regime by .28. I interpret this as a reflection of the much higher likelihood of choosing the automatic regime across all subjects when there is a positive audit rate, due to the deterrence effect, which reduces the difference between Compliers and others. When the audit rate is zero, however, many Compliers choose automatic reporting since they are not going to cheat, whereas the incentive to choose self-reporting is much stronger for the remaining subjects since they can cheat without bearing any risk.

5.f Predicting Stage 3 underreporting using behavior in Stages 1 and 2

Table 7 presents the results of the Tobit regressions in Stage 3, similar to those in Stage 1. A stage 1 regression is also included for comparison, with the sample restricted to those rounds that matched Stage 3 on the penalty rate and tax rate parameters. Coefficients have the same signs and are similar in magnitude to the estimates from Stage 1. The results do not change substantially when only "long form" rounds are included in the regression. In these rounds the reporting burden is essentially the same for auto- and self-reporting. Adding random effects does not alter the coefficients a large amount, as was the case with the probit regressions. It is also worth noting the differences in the estimated effects of the experimental parameters between the two standard Tobit models with different controls. This occurs because these regressions are conditional on choosing the self-reporting regime, and thus the experimental parameters are no longer uncorrelated with each other or the control variables. Overall these results indicate that having a choice of reporting regime does not fundamentally alter the pattern of subjects' response to changes in parameters.

Furthermore, propensity-to-cheat, as measured by both of our indices, increases underreporting, which is in line with expectations. Subjects who were Compliers in Stage 1 still report much less in the self-reporting regime. Willingness-to-pay for burden reduction, conditional on self-reporting, also is associated with more underreporting, but it is difficult to determine whether this effect is actually due to WTP or whether it is simply picking up the effects of correlated omitted variables. One possible explanation for this finding is that subjects with low WTP do not mind filling out the LONG reporting form, and so they may choose the self-reporting regime regardless of how much they plan to earn by cheating. In contrast, those with high WTP have a high cost of filling out the reporting form, and thus only when the benefit of cheating is high, which requires high underreporting, do we see high-WTP subjects choosing the self-reporting regime.

6 Conclusion

I have reported several types of evidence that within this experimental context, the choice between a high-burden, low-transparency and a low-burden, high transparency tax regime causes those with higher propensity-to-cheat on their taxes to self-select into the high-burden regime, even after controlling for cost of reporting. These findings provide evidence that use of a similar mechanism in a natural context could be effective as a separation mechanism, and corroborate the findings of several theoretical studies on the topic. More specifically, we have seen that propensity-to-cheat as measured in Stage 1 does appear to influence the choice of tax regime, and in the expected direction. However, the propensity-to-cheat index only influences choice within rounds where the audit rate is .1. In these rounds there is uncertainty in payoffs in the self-reporting regime when a subject chooses to cheat, but the expected payoffs are still positive, so differences in risk aversion drive differences in reporting behavior. In contrast, when the audit rate is 0, being a Complier significantly affects regime choice, which makes sense because only those with non-economic reasons for not cheating would choose to not cheat in these rounds, after controlling for reporting burden. Surprisingly, the "form type," which is an attempt to vary the cost of choosing the self-reporting regime, does not appear to have any effect on subjects' choice of regime.

In addition to the analysis of regime choice, we have evidence on some of the more specific forces that are driving the difference in cheating behavior among subjects, in the absence of regime choice. Subjects can be usefully separated into Gamers, who sometimes cheat on their income reports and are usually sensitive to the experimental parameters, and Compliers, who never cheat. Among the parameters that we varied experimentally, audit rate had the largest impact on subjects' behavior. This may be due to the fact that the expected value of underreporting income switches from positive to negative as the audit rate changes, while this is not the case for the tax or penalty rates, holding audit rate constant. Moreover, the change from 0 to .1 audit rate is a shift from a certain to an uncertain choice. The signs of the estimated parameter effects were as predicted in the Allingham/Sandmo/Yitzhaki model. However, a large fraction of subjects (the Compliers) never underreported, which cannot be explained by this model. The World Values Survey cheating question sheds some light on the Compliers. Compliers are much more likely to have indicated they believe that cheating on taxes is never justifiable, and lower scores on the WVS index are associated with higher likelihood of being a Complier. The WVS question is designed to capture the non-economic motivation of "tax morale," and thus we can attribute at least a part of the observed Complier behavior to this quality.

Though our experimental context and our population differ in significant ways from natural settings, there is still good reason to believe that our results hold meaningful external validity. To begin with, there is clearly a link between being a Complier and subjects' answers to the WVS question. People who both never cheated in the experiment and said that cheating was never justifiable seem highly unlikely to cheat on taxes in a real-world setting. However, apart from these individuals, there are large discrepancies between our experimental measures of PTC and the WVS question. This fact, in addition to the comments of several subjects over the course of the experiment,[13] makes it seem likely that many of the cheaters in our experiment are not the same people as those who cheat on actual taxes. PTC among these subjects seems more likely to capture general risk aversion than actual propensity to cheat on taxes. But external validity only requires us to assume that within the context of real-world tax reporting, a similar mechanism to the one in this experiment could produce similar results for the set of people with high real-world propensity-to-cheat. Regardless of what is causing a high PTC within certain people, whether low tax morale in a real-world setting or incomplete framing and a "gamer" mentality in our experiment, high PTC should cause separation in the face of a similar regime choice. The same fundamental economic forces are in effect regardless of the specific context that is influencing individual preferences.

There are numerous avenues for future research based on this dataset, including:

—Data were collected on subjects' personality characteristics. Currently this has remained unexamined, and is likely to provide some very interesting results. The relationship between these measures and the WVS question could also be interesting.

—A closer look at reporting burden could also be interesting, and one possibility is using the actual amount of time a subject spent on the reporting screen as an alternative measure to WTP. This may also help shed light on why the form type has no effect on behavior in Stage 3. d —Another area that is very interesting is looking at those subjects in more detail who chose the self-reporting regime but did not underreport. Perhaps these people did not understand the instructions? We have a question asking for subjects' self-assessed understanding of the rules in the

[13] For example, one subject told us "I would never cheat on my taxes but here I cheated every opportunity I got."

experiment that could be used to help answer this and other questions.

—I have begun to pick apart the probit regressions by separating them by audit rate, and I would like to go further in this direction, and extend this type of analysis to other outcomes such as underreporting in Stages 1 and 3.

—Look more closely at individual subjects' behavior in Stage 1, and use it to classify people into types based on apparent motivations for not underreporting and possibly get a measure of risk aversion. This classification system can then be further examined in the context of the regressions separated by experimental state. In particular, some types of subjects would be expected to alter their behavior between the states, based on the expected value of underreporting, while others will not.

—Further investigating how the experimental parameters affect the quality of separation via regime choice in Stage 3 is another interesting direction.

7 References

ALLINGHAM, M. G., and A. SANDMO (1972): "Income Tax Evasion: A Theoretical Analysis," *Journal of Public Economics*, 1, 323-338.

ALM, J. (2012): "Measuring, Explaining, and Controlling Tax Evasion: Lessons from Theory, Experiments, and Field Studies," *International Tax and Public Finance*, 19, 54-77.

ANDREONI, J., B. ERARD, and J. FEINSTEIN (1998): "Tax Compliance," *Journal of Economic Literature*, 36, 818-860.

BECKER, G. S. (1968): "Crime and Punishment: An Economic Approach," *Journal of Political Economy*, 76, 169-217.

BLACK, T., K. BLOOMQUIST, E. EMBLOM, A. JOHNS, A. PLUMLEY, and E. STUK (2012): "Federal Tax Compliance Research: Tax Year 2006 Tax Gap Estimation."

FALKINGER, J., and H. WALTHER (1991): "Separating Small and Big Fish: The Case of Income Tax Evasion," *Journal of Economics*, 54, 55-67.

KALAMBOKIDIS, L., A. TURK, and M. BLUMENTHAL (2012): "Behavioral Experiments of Alternative Reporting Regimes: Transparency Vs. Burden."

RASKOLNIKOV, A. (2009): "Revealing Choices: Using Taxpayer Choice to Target Tax Enforcement," *Columbia Law Review*, 109, 689-754.

SANDMO, A. (2012): "An Evasive Topic: Theorizing About the Hidden Economy," *International Tax and Public Finance*, 19, 1-24.

SLEMROD, J. (2008): "Tax Compliance and Tax Evasion," in *The New Palgrave Dictionary of Economics*: Palgrave Macmillan.

VARIAN, H. R. (1992): *Microeconomic Analysis*. Norton New York.

WOOLDRIDGE, J. (2010): "Econometric Analysis of Cross Section and Panel Data," *MIT Press Books*.

YITZHAKI, S. (1974): "Income Tax Evasion: A Theoretical Analysis," *Journal of Public Economics*, 3, 201-202.

Table 1
Summary of Experiment States

State	Stage	Probability of audit	Penalty rate	Tax rate	Burden reduction fee	Form type
1	1	0	1	0.15		
2	1	0	1	0.3		
3	1	0	2	0.15		
4	1	0	2	0.3		
5	1	0.1	1	0.15		
6	1	0.1	1	0.3		
7	1	0.1	2	0.15		
8	1	0.1	2	0.3		
9	1	0.5	1	0.15		
10	1	0.5	1	0.3		
11	1	0.5	2	0.15		
12	1	0.5	2	0.3		
13	2			0.15	$0.25	Long
14	2			0.15	$0.50	Long
15	2			0.15	$1.00	Long
16	2			0.15	$2.00	Long
17	2			0.15	$4.00	Long
18	3	0	2	0.15		None
19	3	0	2	0.15		Short
20	3	0	2	0.15		Long
21	3	0.1	2	0.15		None
22	3	0.1	2	0.15		Short
23	3	0.1	2	0.15		Long
24	3	0.5	2	0.15		None
25	3	0.5	2	0.15		Short
26	3	0.5	2	0.15		Long

Note: Adapted from Kalambokidis et al. (2012)

Table 2a

Summary of Experiment States, and Compliance by Experiment State, Stage 1

State	Probability of audit	Penalty rate	Tax rate	Percent of subjects non-compliant	Mean fraction under-reported
1	0	1	0.15	62.7	0.51
2	0	1	0.30	63.9	0.52
3	0	2	0.15	63.3	0.51
4	0	2	0.30	65.8	0.52
5	0.1	1	0.15	53.3	0.32
6	0.1	1	0.30	54.2	0.33
7	0.1	2	0.15	48.5	0.27
8	0.1	2	0.30	48.8	0.26
9	0.5	1	0.15	27	0.10
10	0.5	1	0.30	27.3	0.09
11	0.5	2	0.15	23.6	0.08
12	0.5	2	0.30	23.6	0.07

Note: Adapted from Kalambokidis et al. (2012)

Table 2b

Summary of Experiment States, and Burden Reduction by Experiment State, Stage 2

State	Tax rate	Burden reduction fee	Form type	Percent of subjects choosing burden reduction
13	0.15	$0.25	Long	42.1
14	0.15	$0.50	Long	32.1
15	0.15	$1.00	Long	25.2
16	0.15	$2.00	Long	16.4
17	0.15	$4.00	Long	10.3

Note: Adapted from Kalambokidis et al. (2012)

Table 2c

Summary of Experiment States, and Compliance and Regime Choice by Experiment State, Stage 3

State	Tax rate	Probability of audit	Penalty rate	Form type	Percent of subjects choosing self-reporting	Percent of subjects non-compliant[1]	Mean under-reported amount
18	0.15	0	2	None	72.7	78.3	0.50
19	0.15	0	2	Short	72.4	80.7	0.52
20	0.15	0	2	Long	70.9	82.9	0.51
21	0.15	0.1	2	None	57.9	70.2	0.26
22	0.15	0.1	2	Short	59.7	69	0.26
23	0.15	0.1	2	Long	60.6	67	0.26
24	0.15	0.5	2	None	35.2	32.8	0.02
25	0.15	0.5	2	Short	36.7	27.3	0.03
26	0.15	0.5	2	Long	35.8	28	0.03

Notes:

1. For states 18-26, the percent non-compliant is among subjects who chose the self-reporting regime.
2. Adapted from Kalambokidis et al. (2012)

Table 3
Subject Characteristics with Mean Index Values

	N	% of total	Propensity to Cheat: Index 1	Propensity to Cheat: Index 2	Willingness to Pay	WVS cheating index
Female	173	52.4	.46	.28	.15	.30
Male	157	47.6	.48	.32	.18	.33
Age: under 20	61	18.5	.61	.39	.14	.33
Twenties	191	57.9	.47	.32	.14	.35
Thirties	33	10.0	.45	.28	.17	.26
Forties	14	4.2	.30	.15	.24	.17
Fifties	21	6.4	.21	.09	.27	.16
Sixties	10	3.0	.32	.11	.40	.27
Race/ethnicity: White	205	62.1	.46	.30	.17	.25
Asian/Pacific Islander	76	23.0	.56	.35	.13	.51
Black/African American	25	7.6	.40	.19	.16	.28
Hispanic/Latino	9	2.7	.28	.18	.16	.22
American Indian	3	1.0	.58	.25	.60	.11
Other	12	3.6	.37	.23	.08	.39
Student	239	72.4	.50	.32	.15	.33
Not a student	91	27.6	.38	.23	.20	.28
Year in school: Freshman[1]	51	21.3	.60	.37	.15	.34
Sophomore	48	20.1	.51	.32	.16	.33
Junior	42	17.6	.51	.34	.13	.30
Senior	51	21.3	.43	.27	.19	.35
Graduate student	47	19.7	.46	.32	.11	.31
Major: Not economics or related[1]	151	63.2	.50	.31	.16	.30
Economics or related	77	32.2	.48	.34	.13	.38
Undeclared major	11	4.6	.68	.38	.11	.34
Employed	211	63.9	.44	.28	.16	.29
Unemployed	119	36.1	.52	.33	.16	.36
Occupation[2]: Management/professional	43	20.4	.39	.23	.15	.36
Service	43	20.4	.47	.32	.13	.26
Sales and office	42	19.9	.47	.28	.21	.33
Natural resource/construction/maint.	10	4.7	.54	.34	.42	.24
Production/transport./material moving	6	2.8	.42	.28	.09	.22
Other	67	31.8	.42	.29	.14	.26
Business owner or partner	34	10.3	.47	.30	.16	.29
Not a business owner or partner	296	89.7	.43	.26	.18	.32

Notes:

1. Year in school and major are for students only.
2. Occupation is for employed subjects only.
3. Adapted from Kalambokidis et al. (2012)

Table 4

Coefficients from Tobit regressions in Stage 1 with fraction of under-reported income as the dependent variable

	(1)	(2)
Audit rate = 0.1	-0.531	-0.565
	(0.040)***	(0.031)***
Audit rate = 0.5	-1.244	-1.252
	(0.049)***	(0.038)***
tax_rate	0.065	0.103
	(0.227)	(0.167)
penalty_rate	-0.086	-0.086
	(0.034)**	(0.025)***
Income	0.037	0.017
	(0.005)***	(0.005)***
round	0.003	0.011
	(0.005)	(0.004)***
wvs_index	1.190	1.829
	(0.183)***	(0.507)***
wvs_sq	-0.796	-1.493
	(0.202)***	(0.567)***
WTP	-0.077	-0.096
	(0.066)	(0.180)
BusinessOwner	0.076	0.077
	(0.062)	(0.172)
Constant	-0.532	-0.405
	(0.166)***	(0.325)
N	3,950	3,950

* p<0.1; ** p<0.05; *** p<0.01

Notes:

1. Both regressions included demographic controls that are not reported.
2. Column 1 is a standard two-limit Tobit. Column 2 is a random-effects Tobit.
3. "wvs_index" is the subject's response to the question "Is cheating on your taxes, if you have the chance, always justifiable, never justifiable, or something in between?" The original 10-point Likert scale has been rescaled so that 0=never justifiable, 1=always justifiable. "wvs_sq" is wvs_index squared.

Table 5

Correlations between WVS "Cheating" question and the constructed propensity-to-cheat (PTC) indices. Spearman is a specialized correlation coefficient for ordinal-scale variables.

a. All subjects

	Pearson			Spearman		
	wvs_index	PTC1	PTC2	wvs_index	PTC1	PTC2
wvs_index	1.00			1.00		
PTC1	0.24	1.00		0.26	1.00	
PTC2	0.22	0.78	1.00	0.25	0.79	1.00

b. Excluding Compliers

	Pearson			Spearman		
	wvs_index	PTC1	PTC2	wvs_index	PTC1	PTC2
wvs_index	1.00			1.00		
PTC1	0.15	1.00		0.13	1.00	
PTC2	0.13	0.58	1.00	0.12	0.55	1.00

Table 6

Marginal effect from a probit regression of Complier status on WVS index and demographics

	Complier
Age <20	-0.232
	(0.076)***
Age 30-40	0.034
	(0.069)
Age 40-50	0.182
	(0.092)**
Age 50-60	0.260
	(0.075)***
Age 60+	0.129
	(0.112)
wvs_index	-0.220
	(0.073)***
N	330

* $p<0.1$; ** $p<0.05$; *** $p<0.01$

Table 7

Marginal Effects from probit regressions for choice of self-reporting regime in Treatment 3

	(1)	(2)	(3)	(4)
Audit rate = 0.1	-0.131 (0.020)***	-0.132 (0.019)***	-0.706 (0.085)***	-0.706 (0.085)***
Audit rate = 0.5	-0.346 (0.017)***	-0.346 (0.017)***	-1.893 (0.095)***	-1.892 (0.095)***
Income	0.014 (0.002)***	0.012 (0.002)***	0.036 (0.013)***	0.034 (0.013)***
round	-0.002 (0.003)	-0.001 (0.003)	-0.013 (0.013)	-0.012 (0.013)
wvs_index	0.106 (0.030)***	0.090 (0.030)***	0.626 (0.368)*	0.578 (0.363)
WTP	-0.270 (0.031)***	-0.263 (0.030)***	-1.671 (0.378)***	-1.601 (0.373)***
PTC1	0.127 (0.034)***		0.641 (0.426)	
Complier	-0.093 (0.029)***	-0.060 (0.025)**	-0.686 (0.362)*	-0.524 (0.319)
ShortForm	0.010 (0.020)	0.011 (0.019)	0.058 (0.081)	0.059 (0.081)
LongForm	0.002 (0.020)	0.002 (0.019)	0.018 (0.081)	0.019 (0.081)
BusinessOwner	0.063 (0.029)**	0.063 (0.029)**	0.170 (0.350)	0.180 (0.346)
PTC2		0.307 (0.044)***		1.495 (0.536)***
N	2,970	2,970	2,970	2,970

* p<0.1; ** p<0.05; *** p<0.01

Notes:

1. Effects reported are average partial effects for the probability of choosing the self-reporting regime. All regressions include demographic controls.
2. Column 1 is a standard probit model and includes PTC1 as a regressor. Column 2 is a standard probit model and includes PTC2 as a regressor. Columns 3 and 4 replicate columns 1 and 2 using a random-effects probit model.

Table 8

Marginal Effects from probit regressions for choice of self-reporting regime in Treatment 3, by audit rate.

	audit rate=0	audit rate=0.1	audit rate=0.5
Income	0.013	0.019	0.009
	(0.002)***	(0.003)***	(0.003)***
WTP	-0.275	-0.213	-0.309
	(0.042)***	(0.053)***	(0.062)***
PTC1	0.038	0.337	0.078
	(0.052)	(0.058)***	(0.064)
Complier	-0.280	-0.003	0.068
	(0.039)***	(0.049)	(0.053)
ShortForm	-0.005	0.019	0.016
	(0.030)	(0.035)	(0.036)
LongForm	-0.023	0.022	0.006
	(0.030)	(0.035)	(0.037)
N	990	990	990

* p<0.1; ** p<0.05; *** p<0.01

Table 9
Coefficients from Tobit regressions in Treatment 3 with Fraction Underreported as the Dependent Variable

	(1)	(2)	(3)	(4)	(5)
Audit rate = 0.1	-0.596	-0.676	-0.679	-0.659	-0.682
	(0.089)***	(0.065)***	(0.056)***	(0.049)***	(0.045)***
Audit rate = 0.5	-1.377	-1.787	-1.645	-1.487	-1.429
	(0.108)***	(0.093)***	(0.080)***	(0.068)***	(0.063)***
Income	0.048	0.046	0.033	0.016	0.018
	(0.008)***	(0.006)***	(0.005)***	(0.005)***	(0.006)***
NoForm		-0.004	0.011	0.014	-0.014
		(0.069)	(0.059)	(0.052)	(0.045)
ShortForm		0.001	-0.010	-0.011	-0.023
		(0.069)	(0.059)	(0.051)	(0.044)
PTC1			0.953		
			(0.105)***		
WTP			0.442	0.476	0.387
			(0.114)***	(0.098)***	(0.161)**
Complier			-1.171	-0.790	-1.051
			(0.121)***	(0.097)***	(0.166)***
PTC2				2.028	2.148
				(0.117)***	(0.197)***
Constant	-0.533	-0.174	-0.318	-0.146	-0.182
	(0.185)***	(0.169)	(0.163)*	(0.133)	(0.186)
N	989	1,655	1,655	1,655	1,655

* p<0.1; ** p<0.05; *** p<0.01

Notes:

1. All regressions in Stage 3 are conditional on choosing the self-reporting regime.
2. Column 1 is Stage 1 Tobit. Column 2 is Stage 3 Tobit without controls, for comparison with Stage 1. Column 3 is Stage 3 Tobit using PTC1. Column 4 is Stage 3 Tobit using PTC2. Column 5 adds random effects.

Figure 1. Distribution of fraction of income underreported, by audit rate.

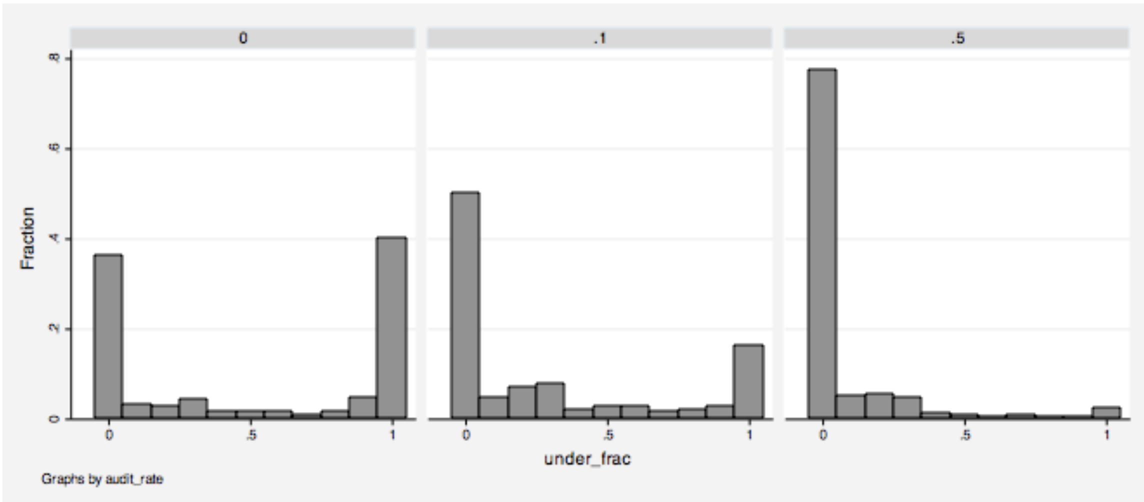


Figure 2. Mean fraction of underreported income by experiment state in Stage 1. Refer to Figure 1 for a summary of the states. Note that states 1-4 have audit rate 0, 5-8 have audit rate 0.1, and 9-12 have audit rate 0.5.

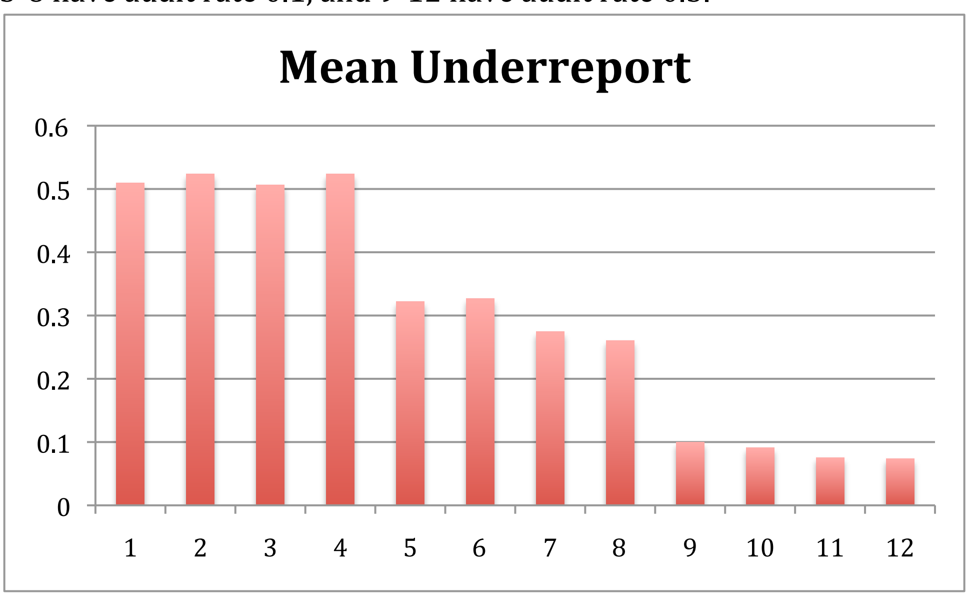


Figure 3a. Graph of overall effect of the WVS cheating question on the average fraction of income underreported, using the full sample. The solid line is from a standard Tobit regression. The dashed line is from a random-effects Tobit.

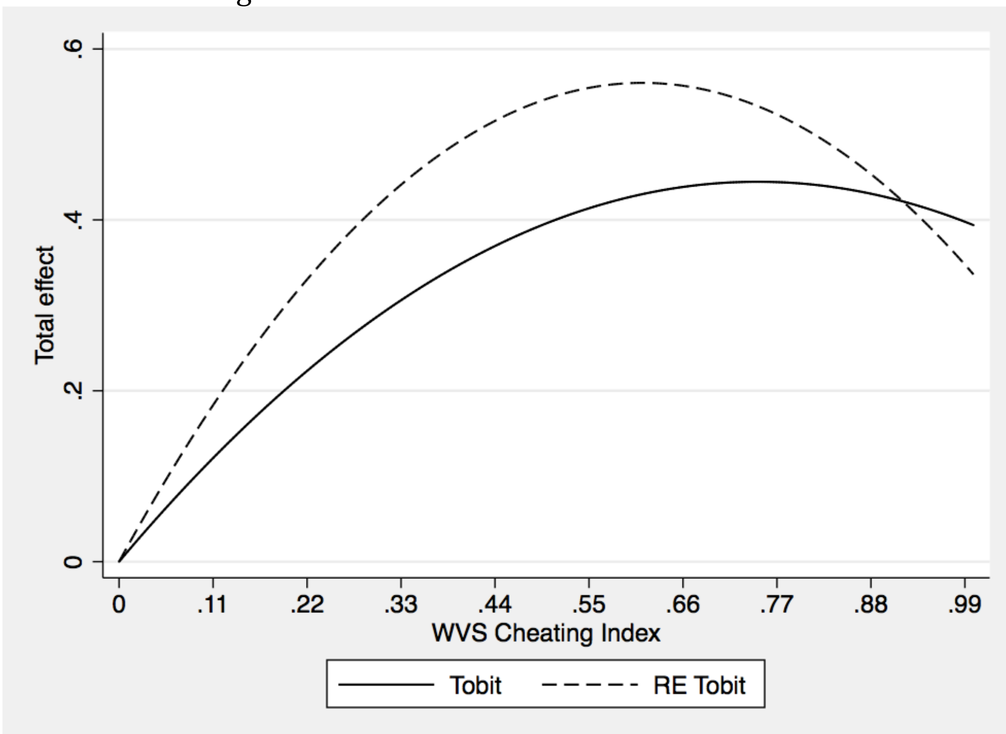


Figure 3b. Graph of overall effect of the WVS cheating question on the average fraction of income underreported, using only "Gamers." The solid line is from a standard Tobit regression. The dashed line is from a random-effects Tobit.

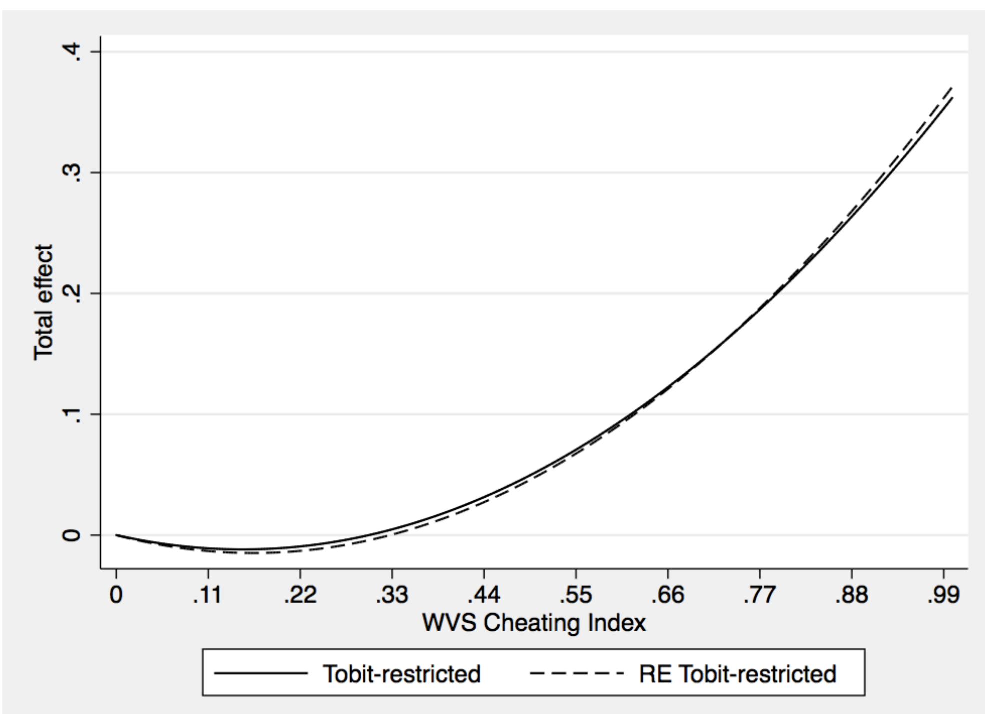


Figure 4. Distribution of WVS cheating index scores, by Complier status.

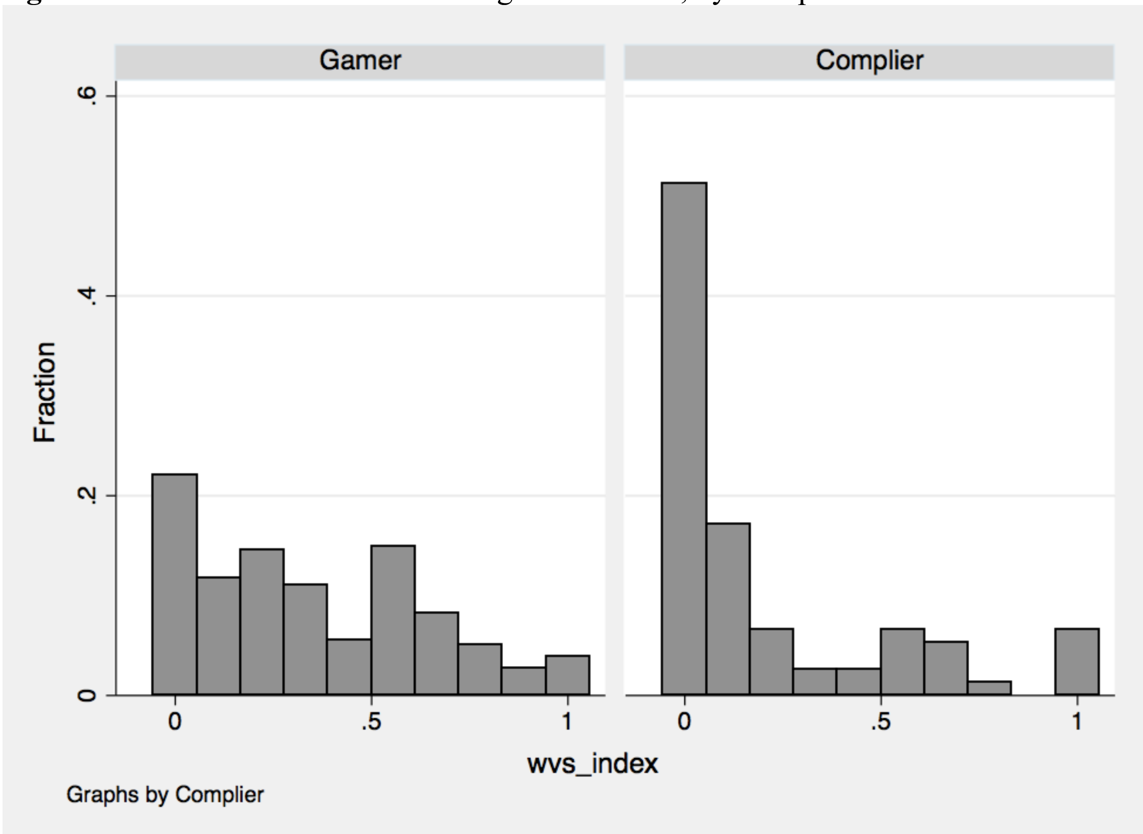
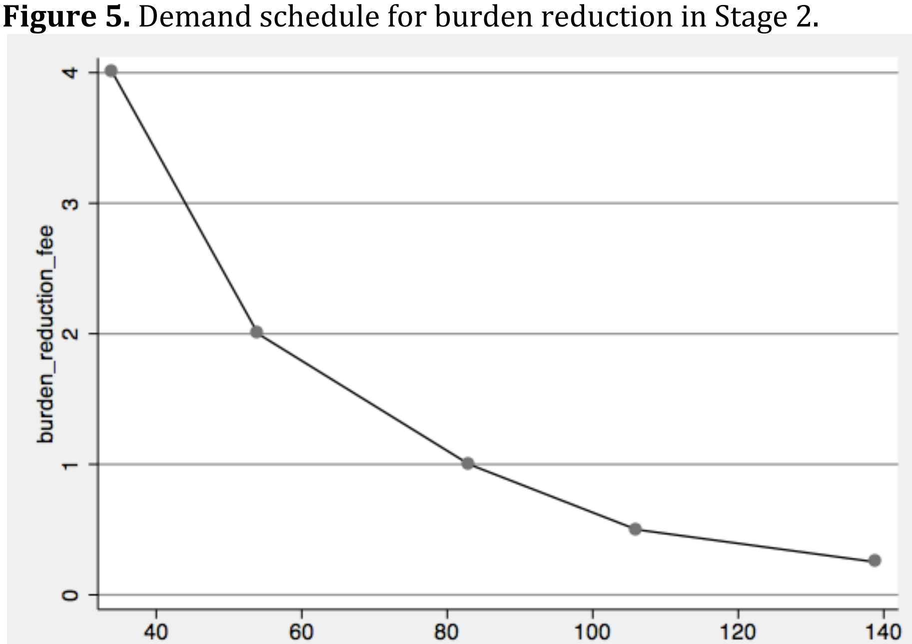


Figure 5. Demand schedule for burden reduction in Stage 2.



Note: Adapted from Kalambokidis et al. (2012)