Predictive Analytics and the Targeting of Audits*

Nigar Hashimzade
Durham University and Tax Administration Research Centre

Gareth D. Myles†
University of Exeter and Tax Administration Research Centre

Matthew D. Rablen
Brunel University and Tax Administration Research Centre

October 28, 2014

Abstract

The literature on audit strategies has focused on random audits or on audits conditioned only on income declaration. In contrast, revenue services employ the tools of predictive analytics to identify taxpayers for audit with a range of indicator variables used for conditioning. The paper explores the compliance and revenue consequences of the use of predictive analytics in a agent-based model that draws upon the behavioral approach to tax compliance. The taxpayers in the model form subjective beliefs about the probability of audit from social interaction, and are guided by a social custom that is developed from meeting other taxpayers. The belief and social custom feed into the occupational choice between employment and two forms of self-employment. It is shown that the use of predictive analytics yields a significant increase in revenue over a random audit strategy by affecting the subjective belief and enhancing the social custom.

*We thank the ESRC for financial support under grant ES/K005944/1.
†Department of Economics, University of Exeter, Exeter EX4 4PU, UK. e-mail: gdmyles@ex.ac.uk

1 Introduction

The standard analysis of tax compliance in Allingham and Sandmo (1972) and Yitzhaki (1974), and much of the literature that has followed, is based on the assumption that taxpayers abide by the axioms of expected utility theory and that audits are random. An exception is the literature on optimal auditing – including Reinganum and Wilde (1985, 1986) and Chander and Wilde (1998) – which characterizes the equilibrium audit strategy as a function of reported income. In practice, the overwhelming majority of audits performed by revenue services are "risk-based" (in which taxpayers are targeted for audit), with only a small fraction of audits performed on a random basis for statistical purposes. Unlike the presumption of the optimal auditing literature, however, the targeting of risk-based audits is not based solely on the income report. Rather, revenue services rely on the experience of case officers reviewing returns and, increasingly, on the basis of predictive analytics which applies statistical tools to the data on a range of taxpayers' characteristics, often in the form of indicator variables (see Cleary, 2011, and the references therein). The expected utility model has also been subject to significant criticism and many alternatives models with behavioral foundations have been proposed.

The paper explores the compliance and revenue consequences of the use of predictive analytics in a agent-based model that draws upon the behavioral approach to tax compliance. We use agent-based modelling because this allows us to explore a richer model than is possible in a theoretical analysis but means we rely on simulation to generate our results. The model is constructed on the foundation of a social network that governs the interaction between taxpayers and the transmission of information between taxpayers. The information consists of attitudes towards compliance (in the form of a social custom) and beliefs about audits (a subjective probability of audit). Taxpayers must make an occupational choice between employment and two forms of self-employment based on their expected income in each occupation. Employment provides a safe income but because of the third-party reporting of income there is no possibility of non-compliance. The two self-employment occupations are risky, but non-compliance is possible. Taxpayers allocate between the occupations on the basis of the expected income from the occupations which accounts for the optimal compliance behavior. Given the different levels of risk in the occupations, taxpayers are divided among occupations on the basis of risk aversion. This results in self-selection of those who will exploit opportunities for non-compliance into occupations where such opportunities arise.

The predictive analytics investigated in the model are based on Tobit and logit regression models using the data revealed in tax returns and data from the outcomes of past audits. The Tobit model targets audits on the basis of predicted evasion level and the logit model on the basis of predicted likelihood of non-compliance. The predictive analytics are implemented by running the model with random audits for an initial period to acquire audit data and then introducing the predictive analytics to predict non-compliance. We consider the outcome when all audits are targeted using predictive analytics and when a combination of targeted and random audits is employed. The "mixed" regimes of targeted and random audit are akin to the random enquiry programmes run alongside targeted audits by the US IRS and the UK HMRC. It is shown that both forms of predictive analytics secure a significant increase in revenue over a random audit strategy.

To give the results validity it is necessary to build the agent-based model on a sound underlying theory of the compliance decision. Our modelling starts from the assumption that taxpayers do not know the audit strategy of the revenue service but must form a belief about the probability of being audited. This is consistent with the idea of behavioral economics that individuals generally do not evaluate risky prospects using the objective probabilities of events but form subjective probabilities (or transform objective probabilities using a weighting function). The subjective probabilities (or, in our terminology, *beliefs*) can differ significantly from the objective probability (Kahneman and Tversky, 1979). There is also empirical (Spicer and Lundstedt, 1976) and experimental (Baldry, 1986) evidence that the individual compliance decision also takes into account social factors such as the perceived extent of evasion in the population. We choose to summarize the range of social factors as the *attitude* of the taxpayer toward compliance. This is essentially identical to the concept of *tax morale* that is prominent in the empirical literature (e.g., Torgler, 2002).

A key feature of our modelling is to make explicit the processes through which the attitude towards compliance and the belief about auditing are formed. Attitudes and beliefs are endogenous and result from the interaction of a taxpayer with other taxpayers and with the revenue service. The importance of interaction makes it necessary to specify the social environment in which the interaction takes place. We do this by employing a social network with a given set of links between taxpayers to govern the flow of information. After each round of audits in the simulation some of the taxpayers who are linked will meet and exchange information. The likelihood of information transmission is greater when the taxpayers are in the same occupation.

The paper is structured as follows. Section 2 describes the separate concepts that are built into the model. Section 3 provides analytical details on how these concepts are implemented. Sections 4 and 5 describe the simulation results under a random audit rule and when the audit rule is informed by predictive analytics. Section 6 concludes.

2 Conceptual Approach

This section describes the elements that constitute the agent-based model. The purpose of the discussion is to relate these elements to the extensive literature on the individual tax compliance decision. The seminal analyses of the compliance decision by Allingham and Sandmo (1972) and Yitzhaki (1974) were built upon the application of expected utility theory. A standard criticism of this model is that it over-predicts the extent of evasion when evaluated using the objective probability of audit[1] which has motivated the application of ideas from behavioral economics The behavioral models of the compliance decision are surveyed in Hashimzade, Myles, and Tran-Nam (2013).

The key elements of our agent-based model is that taxpayers make an *occupational choice* decision prior to the compliance decision. The compliance decision is based on the attitude toward compliance as summarized in a *social custom* and belief about audits captured in a *subjective probability* of audit. The information used to form attitudes and beliefs is transmitted through meetings between taxpayers that a governed by a *social network*. Each of these components is now described in greater detail.

2.1 Occupational choice

Occupational choice determines the possibility of concealing income in different occupations. Income from employment is often subject to a withholding tax and/or third-party reporting to the revenue service. For example, the UK Pay-As-You-Earn system involves income tax being deducted by employers and remitted directly to the revenue service. The prevents evasion by employees except unless there is collusion with the employer so non-compliance is only possible for taxpayers who are self-employed. Occupations also differ in their traditions concerning payment in cash. Those in which cash payment is common provide the greater opportunity for evasion. Occupational choice has not had a prominent role in the literature on tax evasion despite its clear importance. Exceptions to this are Pestieau and Possen (1981) who model occupational choice, and Cowell (1981), Isachsen and Strøm (1980) and Trandel and Snow (1999) who analyze the choice between working in the regular and the informal economies.

Occupational choice is also important for the connection it has with risk aversion. Individuals allocate to occupations on the basis of their ability at that occupation and their attitude to risk. Those who are least risk averse will choose to enter the riskiest occupations. Kanbur (1979) and Black and de Meza (1997) assume employment is safe but self-employment is risky, and address the social efficiency of aggregate risk-taking. Self-employment attracts the least risk-averse taxpayers, who will evade the most when the opportunity arises. Hence, occupational choice has the effect of self-selecting taxpayers who will evade into a situation in which they can evade. This observation should form part of any explanation of why non-compliance can be so significant within specific occupational groups.

Our model includes a choice between employment and two forms of self-employment. Employment is a safe activity that delivers a known income. Self-employment is risky so each taxpayer has to take into account the probability distribution of income when making an occupational choice. One of the self-employment occupations is more risky than the other, in a sense we make precise below[2].

[1] It should be noted that Slemrod (2007) gives good reasons why this claim should be treated with caution.

[2] Individuals differ in their level of skill in the occupations, and skills are one of the deter-

2.2 Social customs

The experiments of Baldry (1986) provide compelling evidence that the evasion decision is not just a simple gamble. This can be rationalized by introducing an additional cost into the evasion decision. These costs can be financial (Bayer, 2006; Lee, 2001) or psychic (Gordon 1989). Psychic costs can arise from fear of detection or concern about the shame of being exposed. The magnitude of the psychic cost can reflect an individual's attitude towards compliance. Attitudes are an important feature of psychological theories of tax compliance (Kirchler, Hoelzl, and Wahl, 2008; Weigel, Hessing, and Elffers, 1987). The psychic cost can also be interpreted as the loss of the payoff from following the social norm of honest tax payment. Adopting this interpretation makes it natural to assume that the size of the loss in payoff is generated by explicit social interaction, and that the size is larger when fewer taxpayers evade (Fortin, Lacroix, and Villeval, 2007;Kim, 2003; Myles and Naylor, 1996; Traxler (2010).

The additional costs have an important role in explaining some features of the tax evasion decision. We model attitudes by including a social custom of honest tax payment in the model so that there is a utility gain (relative to the state with non-compliance) when tax is paid in full. The importance attached to the social custom by each taxpayer is determined by their interaction with other taxpayers within the social network.

2.3 Subjective beliefs

We have already observed that if choice is based on objective probability of audit then the standard model over-predicts the amount of evasion. This has led to the application of choice models based on non-expected utility theories. Non-expected utility models can predict the correct level of evasion for reasonable parameter values. This is because they permit the subjective probability of audit (the weighting on the payoff when audited) to be greater than the objective probability. They also open the possibility of designing compliance policy to manipulate the subjective nature of the decision (Elffers and Hessing, 1997).

It is standard to distinguish between a choice with risk (the decision maker knows the probability distribution of future events) and a choice with uncertainty (the decision maker does not know the probabilities). A first step away from expected utility theory is to consider a choice with risk but to assume the probabilities are distorted into "decisions weights" that enter the expected payoff. Rank dependent expected utility framework (Quiggin, 1981, 1982; Quiggin and Wakker, 1994) uses a particular weighting scheme to transform the objective probability of events into subjective probabilities and has been applied to the evasion decision by Arcand and Graziosi (2005), Bernasconi (1998) and Eide (2001). Prospect theory (Kahneman and Tversky, 1979; Tversky and Kahneman, 1992) also uses a weighting scheme but payoffs are determined by gains and losses relative to a reference point. It has been applied to compliance by

minants of income. This makes it necessary to state the formal details before "more risky" can be explained in full.

al-Nowaihi and Dhami (2007), Bernasconi and Zanardi (2004), Rablen (2010), and Yaniv (1999). Uncertainty has been modelled by assuming the decision maker forms a probability distribution over the possible probability distributions of outcomes ("second-order uncertainty"). This gives rise to the concept of ambiguity (surveyed in Camerer and Weber, 1992) which has been applied by Snow and Warren (2005).

We incorporate these ideas into the model by assuming that each taxpayer forms a subjective belief about the audit probability and explicitly modelling the process for forming beliefs. This allows the model to provide an explanation of how subjective probabilities can endogenously emerge and remain systematically different from the objective probabilities.

2.4 Social network

The illegality of tax evasion and the incentive the revenue service has to conceal its audit strategy imply that taxpayers cannot be fully informed. A natural assumption is that information will not be revealed publicly, but will be transmitted between taxpayers in a position of mutual trust. The social network we adopt is a formalization of this assumption.

The importance of social contacts is supported by empirical evidence on the positive connection between the number of tax evaders a taxpayer knows and the extent of evasion of that taxpayer (De Juan, Lasheras, and Mayo, 1994; Geeroms and Wilmots, 1985; Spicer and Lundstedt, 1976; Wallschutzky, 1984; Webley, Robben, and Morris, 1988). This evidence demonstrates that the compliance decision is not made in isolation but that each taxpayer makes reference to the observed behavior of the society in which they operate.

We capture this social interaction by applying network theory (Goyal, 2009; Jackson, 2004). Networks have previously been used in the analysis of evasion by. Korobow, Johnson, and Axtell (2007) and Franklin (2009). They have also been applied to crime more generally (Glaeser, Sacerdote, and Scheinkman, 1996).

The social network in our model plays two roles. First, it transmits the social custom from one person to another: if two non-evaders meet the importance of the social custom of honest payment is increased for both, but if a non-evader meets an evader then it is reduced for the non-evader and increased for the evader. Second, the network transmits information about audit policy. Since the audit strategy is not public information, taxpayers have to infer it from their own experience and from the receipt of information about the experiences of others. Our simulations are an application of agent-based modelling (Bloomquist, 2004; Tesfatsion, 2006) with agent interaction controlled by the social network.

3 Network Model

In this section we model the formation of attitudes and beliefs as the outcome of social interaction, and opportunities as the outcome of occupational choice.

This is achieved by applying the theory of network formation to track the links between taxpayers and the transmission of attitudes and beliefs, and combining this with agent-based modelling which employs a behavioral approach to describe individual choices.

There are n individuals, indexed $j = 1, ..., n$, interacting repeatedly in discrete time, $t = 1, ..., T$. Each individual, j, at time t is characterized by a vector of characteristics

$$\left(w_j, \rho_j, s_j^1, s_j^2, z_j; p_{j,t}^0, p_{j,t}^1, p_{j,t}^2, \chi_{j,t}\right). \tag{1}$$

At the start of the simulation values for all characteristics are randomly assigned to each taxpayer by making draws from independent distributions. The first five characteristics remain constant throughout the simulation. These characteristics are w_j, the wage in employment (occupation 0); ρ_j, the coefficient of relative risk aversion; s_j^α, the skill in self-employed occupation α, $\alpha = 1, 2$; and z_j, the payoff from following the social custom. The remaining four characteristics are updated through interaction with the revenue service and with other taxpayers in the social network so are indexed by time, t. These are: $p_{j,t}^\alpha$, the perceived (subjective) probability of audit in occupation α, $\alpha = 0, 1, 2$, and $\chi_{j,t}$, the weight attached to payoff from the social custom.

We now describe how these variables enter into the choice problem of a taxpayer and how the subjective probability and weight attached to social custom are updated.

In each period, t, every individual chooses their preferred occupation[3] and, once income is realized, the optimal level of evasion. Individual j has a choice between employment or entering one of the two self-employment occupations. If employment is chosen the wage, w_j, is obtained with certainty. The self-employment opportunities are represented as risky "projects". The outcome of self-employment for individual j in occupation α at time t is given by $s_j^\alpha y_{j,t}^\alpha$ where $y_{j,t}^\alpha$ is a random draw at time t from the probability distribution function $F^\alpha(\cdot)$. The choice of occupation is taken on the basis of $F^\alpha(\cdot)$ but the choice of evasion is made after the realization of $y_{j,t}^\alpha$. It is assumed that $E\left(y^1\right) < E\left(y^2\right)$ and $Var\left(y^1\right) < Var\left(y^2\right)$, so if $s_j^1 = s_j^2$ occupation 2 is riskier than occupation 1 but offers a higher expected income. Both self-employment occupations are riskier than employment, in the sense that for each agent the wage in employment is certain, *i.e.* $Var\left(w_j\right) = 0$.

It is not possible to evade tax in employment because incomes are subject to third-party reporting or to a withholding tax. Evasion only becomes possible when self-employment is chosen. In occupation α taxpayer j has belief at time t that the probability of evasion being detected is $p_{j,t}^\alpha$. The belief about the probability of detection is updated through the experience of the taxpayer with audits and through the exchange of information when meeting other taxpayers. The attitude of taxpayer j toward evasion is summarized in $\chi_{j,t}$, the weight given to the social custom. This attitude is also updated through meetings with

[3] It may seem unrealistic to have an occupational choice in every period but in the simulations only a very small number of taxpayers actually change occupation in any period.

other taxpayers. We describe the processes for updating attitudes and beliefs in detail after discussing the choice of occupation for given attitudes and beliefs.

The choice of occupation and the choice to evade tax involve risk. Taxpayer j has a (constant) degree of relative risk aversion measured by the risk aversion parameter, ρ_j. The taxpayer chooses occupation and evasion level at time t to maximize subjective expected utility given beliefs $\{p^{\alpha}_{j,t}\}$. For analytical tractability, we assume throughout a CRRA form for utility:

$$U\left(Y_j\right) = \frac{Y_j^{1-\rho_j} - 1}{1-\rho_j}. \tag{2}$$

The attitude toward evasion determines the utility value of following the social custom of honest tax payment. The payoff from the social custom is given by z_j and the individual weight, or the importance, assigned to this payoff by the taxpayer is determined by $\chi_{j,t}$. Hence, compliance with tax payment at time t generates an additional utility from the social custom of $\chi_{j,t} z_j$.

In employment there is no opportunity for evasion so the taxpayer obtains a payoff given by

$$V^0 = \frac{\left[(1-\tau)w_j\right]^{1-\rho_j} - 1}{1-\rho_j} + \chi_{j,t} z_j, \tag{3}$$

where τ is the constant marginal tax rate. The possibility of tax evasion makes the choice of self-employment a compound lottery: the outcome of the project is random, as is the outcome of choosing to evade.

Define the expected payoff from the optimal choice of evasion in self-employment occupation α for a given realization $y^{\alpha}_{j,t}$ as

$$\begin{aligned} V^{\alpha}_e\left(y^{\alpha}_{j,t}\right) &= \max_{E^i_j \in \left[0, s^i_j y^i_{j,t}\right]} \left\{ p^{\alpha}_{j,t} U\left(s^{\alpha}_j y^{\alpha}_{j,t} - f\tau E^{\alpha}_j\right) \right. \\ &\qquad \left. + (1-p^{\alpha}_{j,t}) U\left(s^{\alpha}_j y^{\alpha}_{j,t} + \tau E^{\alpha}_j\right) + \chi_{j,t} z_j \mathbf{1}_{\left[E^i_j = 0\right]} \right\}, \end{aligned} \tag{4}$$

where $f > 1$ is the fine levied on unpaid tax if evasion is detected. The term $\mathbf{1}_{[A]}$ is an indicator function that takes the value of one if A is true and zero otherwise: the payoff from the social custom is obtained only if tax is paid in full. The level of evasion will be a function $E^{\alpha}_j\left(y^{\alpha}_{j,t}\right)$ of the realized income $y^{\alpha}_{j,t}$ in occupation α. The expected payoff from the compound lottery describing occupation α is then

$$V^{\alpha} = \int_{Y^{\alpha}} V^{\alpha}_e\left(y\right) dF^{\alpha}\left(y\right). \tag{5}$$

The choice of occupation is made by comparing the utility levels from employment and from self-employment. Hence, the chosen occupation is given by selecting the maximum of $\left\{V^0, V^1, V^2\right\}$.

After self-employment occupation α is chosen at time t an outcome $\tilde{y}^{\alpha}_{j,t}$ is realized according to the probability distribution function $F^{\alpha}\left(\cdot\right)$. Given the outcome, the optimal evasion decision is implemented, as described above. Denote the level of evasion that is realized by $\tilde{E}^{\alpha}_{j,t} = E^{\alpha}_{j,t}\left(\tilde{y}^{\alpha}_{j,t}\right)$. Tax returns are submitted, and a number of those in self-employment occupations are then audited, according to a process chosen by the revenue service. If evasion is discovered, unpaid tax is reclaimed and the fine at rate f is imposed on unpaid tax.

The social network is modelled as a set of bidirectional links described by an $n \times n$ symmetric matrix of zeros and ones. For example, in the network described by matrix A

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}. \tag{6}$$

the first row, representing the links of individual 1, has a single 1 in column 2 which means that 1 is linked to 2. There is a corresponding 1 in the first column in the second row representing the links of individual 2 with 1. That is, the element in row i and column j of matrix A is defined as

$$A_{ij} = \begin{cases} 1 & \text{if } i \text{ and } j \text{ are linked in the network,} \\ 0 & \text{otherwise.} \end{cases}$$

In the simulation, the matrix is created at the outset and does not change.[4] The network determines who may meet whom to exchange information. In each period a random selection of meetings occur described by a matrix C^t of zeros and ones which is randomly selected every period. Individuals i and j meet during period t if $A_{ij}C^t_{ij} = 1$. At a meeting of i and j there is a probability that information is exchanged about the subjective probability of audit and whether the taxpayer was compliant in that period. The probability of information exchange depends on the occupational groups to which i and j belong; the probability is highest when they are in the same occupation. Let i be engaged in occupation α and j in occupation β. The probabilities of information exchange occurring at a meeting are given by $q^{\alpha\beta}$ where $\alpha, \beta = 0, 1, 2$.

Recall that individuals hold beliefs about the probability of being audited in each of the three occupations. We assume there are two ways in which beliefs are updated. Consider taxpayer j who has worked in occupation α in period t. After submission of the tax return, the taxpayer may or may not be audited. On the basis of the outcome the belief about the audit probability, $p^{\alpha}_{j,t}$, in that occupation is then adjusted. The beliefs about the audit probability in the other two occupations, $p^{\beta}_{j,t}$, $\beta \neq \alpha$, remain unchanged at this stage. Following this, the taxpayer may meet with a contact in the network. Let the meeting be with a taxpayer who is engaged in occupation β. At the meeting information is exchanged with probability $q_{\alpha\beta}$. This information is then used to update the belief about the audit probability, $p^{\beta}_{j,t}$, in the occupation β of the contact. The other two beliefs remain unchanged since no information is communicated.

[4] Here the network is fixed, but the probabilities of information exchange between the linked individuals change if they switch occupations, as described below. Another possibility would be to have the network itself revised as a consequence of chosen actions, i.e. agents in different occupations belonging to different social networks.

The choice of occupation in period $t+1$ is made on the basis of the beliefs $\{p^0_{j,t}, p^1_{j,t}, p^2_{j,t}\}$ updated after the audits and the information exchange. Two different processes for the updating of subjective beliefs following an audit are considered. As studies have reliably demonstrated important deviations from Bayesian inference (e.g. Grether, 1980), we allow for non-Bayesian updating. The first process, which is qualitatively similar to a Bayesian process, is to assume that individuals feel marked as targets if they are audited, so that one audit is believed likely to be followed by another. We term this the *target effect.* In contrast, those not audited in a period believe they are less likely to be audited in the next period. Formally, if audited in period t, an individual's belief about being audited in the next period is raised to probability P, otherwise it decays. The updating rule for the subjective probability is therefore

$$\begin{aligned} \tilde{p}^{\alpha}_{j,t+1} &= AX_{j,t}P + (1 - A_{j,t})\,\delta p^{\alpha}_{j,t}, \;\; \delta \in [0,1]\,, \;\; P \in [0,1] \\ \tilde{p}^{\beta}_{j,t+1} &= p^{\beta}_{j,t+1}, \;\; \beta \neq \alpha. \end{aligned} \tag{7}$$

where $A_{j,t} = 1$ if taxpayer j was audited in period t and $A_{j,t} = 0$ otherwise. This can also be written as

$$\tilde{p}^{\alpha}_{j,t+1} = \begin{cases} P \in [0,1] & \text{if audited at } t, \\ \delta p^{\alpha}_{j,t+1}, \;\; d \in [0,1] & \text{otherwise.} \end{cases} \tag{8}$$

We refer to the case of $P = 1$ as the maximal target effect.

As an alternative, we have also considered a second process that captures the *bomb-crater* effect documented experimentally by Guala and Mittone (2005), Kastlunger, Kirchler, Mittone, and Pitters (2009), Maciejovsky, Kirchler, and Schwarzenberger (2007) and Mittone (2006). In this process a taxpayer who has been audited in one period believes that they are less likely to be audited in the next (or not audited at all), but the belief slowly rises over time. The process is therefore described by

$$\tilde{p}^{\alpha}_{j,t+1} = \begin{cases} P \in [0,1] & \text{if audited at } t, \\ p^{\alpha}_{j,t} + \delta\left(1 - p^{\alpha}_{j,t+1}\right), \;\; \delta \in [0,1] & \text{otherwise,} \end{cases} \tag{9}$$

with $P = 0$ being the maximal bomb-crater effect.

After the audit process is completed the taxpayer may meet with a contact. The information that may (or may not) be exchanged at a meeting includes the subjective probabilities and whether or not the agents were audited. If taxpayer j in occupation α meets individual i who works in occupation β *and if information exchange occurs* at the meeting, the subjective probability is updated according to the rule

$$\begin{aligned} p^{\beta}_{j,t+1} &= \mu \tilde{p}^{\alpha}_{j,t} + (1 - \mu)\,\tilde{p}^{\beta}_{i,t}, \;\; \text{if } \alpha = \beta, \\ p^{\gamma}_{j,t+1} &= \tilde{p}^{\gamma}_{j,t}, \quad \text{if } \alpha \neq \beta. \end{aligned} \tag{10}$$

The importance assigned to the social custom is also determined by interaction in the social network. The weight, $\chi_{j,t}$, is updated in period t *if information* *exchange occurs* between j and some other taxpayer in that period. Assume individual j meets individual i at time t and information exchange takes place. The updating process is described by

$$\chi_{j,t+1} = \frac{1}{M(j)+1}\left[\chi_{j,t}M(j) + \mathbf{1}_{\left[\tilde{E}_i^{\alpha}=0\right]}\right], \tag{11}$$

where $M(j)$ is the number of previous meetings for j at which information was exchanged and $\tilde{E}_i^{\alpha}$ is the level of evasion of i. Hence, $\chi_{j,t+1} > \chi_{j,t}$ if information is exchanged with an compliant taxpayer, and $\chi_{j,t+1} < \chi_{j,t}$ if information is exchanged with an evader.

4 Baseline Simulations

We first conduct a simulation of the network model described above under the assumption of random audits to obtain a baseline outcome. This allow an investigation of the nature of the equilibrium and the consequences of the alternative updating rules for beliefs.

The results we present assume the target effect for audits as specified in equation (8). The results for the bomb-crater model differ only in the pattern of compliance after audit, as described below; the outcomes for the revenues are not qualitatively different from those under the target effect. A complete set of results for the bomb-crater model are available from the authors upon request. The parameter values and the distributions for the random variables that remain constant across the simulations are given in the Appendix. As seems realistic, we set the parameter values such that, on average, the payoff from self-employment will exceed that from employment.[5]

We assume that earnings in occupation α, $\alpha = 0, 1, 2$, are drawn from lognormal distribution, $\log\mathcal{N}(\mu_\alpha, \sigma_\alpha^2)$, and that skills in self-employment are drawn from $\frac{1}{1-\gamma U}$, where U is a uniform $[0,1]$ random variable, and $\gamma \in (0,1)$ is a constant parameter. Each individual knows their wage in employment, $y_{j,t}^0 = w_j$, skill, s_j^i, in the self-employment occupations $i = 1, 2.$, and the distribution of outcomes, $F^{\alpha}(\cdot)$, in the self-employed occupations. Each individual is randomly assigned a vector of subjective beliefs, $\left\{p_{j,0}^0, p_{j,0}^1, p_{j,0}^2\right\}$, and the level of importance of social custom, $\chi_{j,0}$, at time 0, from a uniform $[0,1]$ distribution. The probability of a random audit for all self-employed is 0.05; the employed are not audited.

The results of the simulations for random audits are illustrated in Figure 1 and provide the baseline against which to assess the effects of predictive analytics. Two simulations were run that differed in the in the probability of information exchange between different groups. The first simulation (denoted

[5]The value of the social custom z is measured in units of utility. Therefore, although z appears constrained to take very small values, these values are comensurate with the values taken by the utility function in (2). Thus, with given parameterisation, a true report increases the utility of an "average" individual by about 10 per cent.

Foc) used *focussed* information transmission. That is, at a meeting information is exchanged with positive probability only between linked agents in the same occupation. The second simulation (denoted Diff) used *diffused* information transmission. In this case there is a positive probability that a meeting between linked taxpayers in dissimilar occupations results in information exchange and the probability of information exchange at meetings between members of the same occupation is reduced compared to that under the focussed information transmission. The outcome reported in Figure 1 is for diffused information exchange. The summary statisitcs in Tables 1 and 3 report results for both focussed and diffused information transmission.

The central message from Figure 1 is that sub-groups of the population (the occupational groupings) can endogenously form different attitudes to compliance. As expected, the operation of self-selection sorts those who are most willing to accept risk into the riskiest occupation (self-employment 2). The seld-employment gives them the opportunity to evade, and they make use of this opportunity to become the least compliant group. The updating process for beliefs and the transmission of information around the social network result in the subjective probability of audit being above the true value for the self-employed. The self-employed groups hold similar beliefs, which are distinctly different from those of the employed. The non-zero belief for the employed reflects their learning about audits from meeting with self-employed. The operation of the social custom results in the employed placing a high weight on honesty. Taxpayers in the two self-employment occupations have a much lower weight but this is not significantly different between occupations. In contrast, with focussed information exchange a significant difference in beliefs and honesty weights can emerge between self-employment occupations.

INSERT FIGURE 1 HERE

The means and the standard deviations for three of the endogenous variables calculated over the last 80 periods in the simulations are given in Table 1. The effect of self-selection into occupations is clearly seen in the mean level of risk aversion in each occupation. Risk aversion is significantly lower in occupation 2 than in occupation 1, and both are lower than in employment. The table also confirms that the subjection belief is above the true value of 0.05 for those in self-employment and that under focussed information exchange it can differ between self-employment occuapation. Compliance of those in employment is equal to 1 by definition. For both forms of information exchange compliance is lower for taxpayers in occupation 2.

Table 2 shows the population means and the standard deviations for wage in employment and skills in self-employment. Table **??** shows the means and standard deviations over occupational groups for the wage in employment and skills in self-employment. The columns are the chosen occupation and the rows are the skills in the different occupations. Hence, the average skill in occupation

	Occupation 0 Employment		Occupation 1 Self-employment		Occupation 2 Self-employment	
	Foc	Diff	Foc	Diff	Foc	Diff
Risk Aversion	3.448 (0.007)	3.586 (0.004)	2.604 (0.003)	2.671 (0.012)	2.337 (0.003)	2.300 (0.011)
Belief	0.000 (0.007)	0.022 (0.004)	0.171 (0.011)	0.191 (0.013)	0.176 (0.013)	0.190 (0.012)
Compliance	1 (0)	1 (0)	0.888 (0.008)	0.900 (0.006)	0.834 (0.012)	0.823 (0.014)

Table 1: Descriptive statistics: random audits, focused and diffused information exchange.

Occupation 0 Employment	Occupation 1 Self-employment	Occupation 2 Self-employment
Population Wage	Population Skill	Population Skill
7.00 (3.162)	1.848 (0.765)	1.848 (0.765)

Table 2: Population wages and skills

1 of a taxpayer who has chosen occupation 2 is 1.440. It can be seen that, once the individuals self-select into a particular occupation, the average productivity in each occupation (wage in safe employment and skill in self-employment on the main diagonal) are above the corresponding average for the population and the standard deviations are reduced.

5 Random Audits and Predictive Analytics

The role of *predictive analytics* is to identify the best audit targets, in terms of the expected level or the expected likelihood of non-compliance. Predictive analytics are used by revenue services including the IRS and HMRC. The IRS, for instance, uses information from its random audit program to design discrim-

	Occupation 0 Employment		Occupation 1 Self-employment		Occupation 2 Self-employment	
	Foc	Diff	Foc	Diff	Foc	Diff
Wage	9.775 (0.002)	10.740 (0.013)	5.619 (0.003)	5.610 (0.012)	5.820 (0.007)	5.729 (0.016)
Skill in SE1	1.340 (0.001)	1.396 (0.000)	2.379 (0.002)	2.318 (0.005)	1.440 (0.001)	1.450 (0.006)
Skill in SE2	1.366 (0.000)	1.488 (0.000)	1.519 (0.001)	1.505 (0.001)	2.419 (0.003)	2.419 (0.008)

Table 3: Wages and skills across occupations

inant functions (DIF) that are used to assign each tax return with a score – called the DIF score – for the likelihood that it contains some irregularities or evasion. Various methods are used for identifying targets for risk-based audits. We wish to explore the effects of predictive analytics on behavior and to examine the extent to which they can improve on other audit strategies.

The method we employ to conduct the analysis is to embed predictive analytics within the agent-based model. We then compare the outcome with predictive analytics based on tax return data to that of random audits. The two forms of predictive analytics we investigate involve econometric analysis for predicting the level of non-compliance (level-targeting) and the probability of non-compliance (rate-targeting) by each taxpayer on the basis of the information provided on the tax return and past audits.

This process is implemented in the simulation by using random audits (with each self-employed individual facing a 0.05 probability of being audited) for the first 100 periods to eliminate the effect of the initial conditions and to accumulate audit data. The data from the last 5 random audits is collected and, after the outcome in periods 96 to 100 is known, is used to estimate a regression equation with the dependent variable being the amount of under-reported income[6] in the first exercise and a binary variable taking the value of one if an individual under-reported their income and zero if reported truthfully in the second exercise. The explanatory variables are the observed characteristics and the audit history of an agent; in our model these are termed *occupation*, *declaration*, and *previous audit.*[7] The estimated equation is used to predict non-compliance given information collected in period 101. From this point onward, the estimated regression equation in period t is then used to predict non-compliance using tax return and audit history data in period $t+1$; the audit outcomes in period $t+1$ are added to the data set, and the regression analysis is used again to predict non-compliance in $t+2$, and so on.

The simulations are used to compare a *targeted regime* in which audits are based entirely on predictive analytics with *mixed regime*s in which a combination of predictive analytics and random audits are used. In each period in which predictive analytics are used the agents are sorted according to either their predicted level of evasion (*largest evaders*) or according to their predicted probability of evasion (*most likely evaders*). In the targeted regime the top five per cent of taxpayers by predicted non-compliance are audited (this means that the number of targeted audits is equal to the average number of random audits, so that the audit costs are, on average, equal between these two strategies). In the mixed regime with 50 percent targeted the top 2.5 percent of taxpayers by predicted non-compliance are audited, and the rest of the self-employed taxpayers are randomly audited, with each facing a 0.025 probability of being selected. The mixed regimes with 75 percent targeted and 25 percent targeted are defined

[6] Alternatively, instead of under-reported income one can use unpaid tax as the variable of interest. In our model these approaches are equivalent because of the flat tax schedule.

[7] In the regression "previous audit" is a binary variable, one if the individual was audited in the previous period and zero otherwise. This specification can be easily extended to include a longer audit history and/or previous audit outcomes.

Variable	ME (avg data)	ME (indiv avg)
Declared Income	-1.3586	-4.4411
Previous audit	-1.1952	-3.9069
Self-employment 1	-0.2424	-2.7047

Table 4: Marginal effects in the evasion level equation.

similarly. The mixed regimes are used to explore the possibility that the agents learn about the audit patterns by exchanging information in their networks: if the audits concentrate on one occupation, individuals in the other occupation may comply less, knowing that they are likely to get away with it.

5.1 Targeting largest evaders

Since expected non-compliance is bounded below by zero (we do not allow for mistaken over-declaration) a Tobit (censored) regression is employed to predict the expected level of evasion. Table 4 shows the marginal effects of explanatory variables upon the predicted level of under-reported income. In particular, individuals with higher declaration, those audited in the previous period, and those in self-employed occupation 1 are predicted to evade less tax. The latter implies that, *ceteris paribus*, the operational audits will tend to focus on individuals in occupation 2. The lower evasion level of the previously audited taxpayers reflects the assumption of the target effect, as opposed to the bomb-crater effect (under the bomb-crater model the previous audit has a positive effect on evasion, as one would expect in this case).

The simulation results for targeted audits are shown in Figure 2. The effect upon revenue can be seen by the increase after period 50 in the first panel. The higher level of revenue is sustained for the remaining periods of the simulation. With random audits the subjective beliefs of the two self-employment occupations are the same. Once targeted audits are imposed the belief in occupation 2 is sustained at the same level as for random auditing but that in occupation 1 falls. Compliance rises significantly in both occupations which is driven by an increases in the honesty weight. The increase in compliance and in the honest weight are mutually supporting through the dynamic process for the social custom.

INSERT FIGURE 2 HERE

Figure 3 demonstrates the effect of predictive analytics on revenues. The figure shows the empirical cumulative distribution function (cdf) of revenue for the the targeted regime and three mixed regimes with different proportions of targeted audits and the cdf for random audits. It can be seen that as the proportion of targeted audits increases the empirical cdf for the mixed regime moves smoothly from that of random audits to that of targeted audits[8]. The figure shows that targeting first-order stochastically dominates the mixed regimes.

INSERT FIGURE 3 HERE

The details of the simulation outcome for the mixed regime with 25 percent of audits targeted are shown in Figure 4. When this is contrasted to the outcome in Figure 2 some significant differences can be seen. There is an increase in revenue after the mixed audits are implemented but the increase is less marked. The fall in the subjective belief of taxpayers in occupation 1 after the introduction of predictive analytics is accompanied by a slight fall in the belief in occupations 2. As a consequence the level of compliance in occupation 1 is reduced relative to random audits but that of occupation 2 is increased. Both are much lower than with targeted audits. The mixed regime also fails to increase the honesty weight in contrast to the targeted regime for which it is significantly increased. The cdf has shown that this mixed regime raises welfare over the random regime because the 25 percent of audits that are targeted are successful. The detailed figures show that this is not accompanied by a general improvement in the compliance environment.

INSERT FIGURE 4 HERE

These results show clearly that the use of predictive analytics increases compliance and results in higher tax and fine revenues. The increase in compliance raises the chance of a meeting with a compliant taxpayer and thus leads to a steady increase in the importance of social custom of honest reporting when predictive analytics are in operation. Compliance is not uniformly increased in occupational groups when random audits are included because of the reduction in focus on the least compliant occupation 2. The simulation therefore demonstrates that a policy of targeted audits outperforms random and mixed auditing.

5.2 Targeting most likely evaders

In the previous sub-section the operational audits were targeted at the individuals with the higher predicted level of non-compliance, or undeclared income. An alternative strategy is to target those predicted most likely to evade, or those with the highest evasion score. The evasion score is similar to the credit score used by credit-rating agencies, and can be calculated by estimating a logit or probit regression. The evasion score is assigned the value of one if the individual

[8] Each panel is for a different draw of the basic parameters so the values are not exactly comparable. The two CDFs in each panel are comparable.

Variable	ME (avg data)	ME (indiv avg)
Declared Income	-0.5673	-0.0027
Previous audit	-0.7540	-0.0035
Self-employment 1	0.7882	0.0037

Table 5: Marginal effects in evasion score equation.

evaded tax and zero if declared truthfully. For the explanatory variables we use the same observed characteristics of the agents and their audit histories as in the previous exercise.

The predicted values of the evasion score from the regression have the interpretation of the predicted probabilities that an individual will under-report their income. In every period the individuals are ranked according to their evasion score, and those with the highest score are audited. Again, the number of audits is equal to the average of the number of random audits, in order to equalize the audit costs across strategies. We compare random audits with targeted audits and with a mix of random and targeted audits. We present the results obtained from the logit regression; the results of the probit regression are very similar and are available upon request.

Table 5 shows the marginal effects of explanatory variables on the predicted probability of evasion. In particular, individuals with higher declared income, those audited in the previous period, and those in self-employment 2 are less likely to evade. These results are similar to the predictions of the model for level-targeting except for the effect of occupation on risk. The explanation for the difference can be seen by considering the behaviour of beliefs and compliance when the most likely evaders are targeted.

Figure 5 displays the outcome of the simulation when the most likely evaders are targeted but 50 percent of audits are random. It can be seen that there are three distinct regimes. The outcome with random audits has the same features as for targeting the largest evaders. Once predictive analytics are introduced the level of complaince is initially increased for taxpayers in occupation 2 but reduced for those in occupation. During this period audits are focusing on occupation 2. After 100 periods there is a second change in regime. The focus of audits becomes occupation 1 and thiere compliance rises significantly while that of occupation 2 falls. These changes are mirrored in the behavior of the subjective belief. This shows that the predictive analytics can switch focus when applied to the most likely evaders. This feature has not been observed in any of the simulation that target the largest evaders.

INSERT FIGURE 5 HERE

The revenues generated by targeting the largest evaders and the most likely evaders are very close, which suggests that these strategies can be equally successful in closing the tax gap.

6 Conclusions

The optimal design of audit strategy is important for revenue services, whose aim is to design policy instruments to reduce the tax gap (the difference between anticipated and actual tax revenue). In this paper we analyze two alternative strategies that use the concept of predictive analytics: targeting the largest evaders and targeting the most likely evaders. We do this is in a rich network model in which taxpayers are heterogeneous in risk, beliefs, and attitude towards compliance, and in which agents may self-select into different occupational groups. In this model, attitudes and beliefs endogenously emerge that differ across sub-groups of the population and behavior is different across occupational groups, and this is reinforced by the development of group-specific attitudes and beliefs. Given this behavior, the tax authority may wish to condition its audit strategy not only on reported income, but also on occupation.

What does our model suggest for the optimal strategy of a tax authority? On the one hand, given the objective of maximizing revenues, targeting the level, or the value, of evasion appears to be more important. On the other hand, the "strike rate", or the proportion of audits that reveal evasion, is also important, if the tax authority wants to reduce the burden on the compliant population (James, 2011). Our results imply that the two strategies have the same quantitative effect on the revenues; furthermore, the rates of compliance in population and in each occupation are not statistically different between the two strategies. The robustness of this conclusion to the selection of model parameters is, however, an issue that could be addressed in future research. Another strategy that would be interesting to investigate is the "light touch", where either random or targeted audit can reveal only a fraction of concealed income but at a lower cost to the tax authority. The "light touch" audits allow a wider coverage of population, thereby increasing subjective beliefs and improving compliance; however, partial detection increases expected payoff from evasion and encourages non-compliance (Rablen, in press). This trade-off, along with the cost considerations can lead to the selection of an optimal mix of audits.

References

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

al-Nowaihi, A., and Dhami, S. (2007). Why do people pay taxes: Expected utility theory versus prospect theory. *Journal of Economic Behavior and Organization*, *64*, 171–192.

Arcand, J.-L., and Graziosi, G. R. (2005). Tax compliance and rank dependent expected utility. *The Geneva Risk and Insurance Review*, *30*, 57–69.

Ashby, J. S., Webley, P., and Haslam, A. S. (2009). The role of occupational taxpaying cultures in taxpaying behaviour and attitudes. *Journal of Economic Psychology*, *30*, 216–227.

Baldry, J. C. (1986). Tax evasion is not a gamble. *Economics Letters*, *22*, 333–335.

Bernasconi, M. (1998). Tax evasion and orders of risk aversion. *Journal of Public Economics*, *67*, 123–134.

Bernasconi, M., and Zanardi, A. (2004). Tax evasion, tax rates and reference dependence. *FinanzArchiv*, *60*, 422–445.

Black, J., and de Meza, D. (1997). Everyone may benefit from subsidising entry to risky occupations. *Journal of Public Economics*, *66*, 409–424.

Bloomquist, K. M. (2004). Multi-agent based simulations of the deterrent effects of taxpayer audits. In Kalambokidis, L. (Ed.), *Proceedings of the 97th Annual Conference on Taxation* (pp. 159–173). Washington, DC: National Tax Association.

Camerer, C., and Weber, M. (1992). Recent development in modelling preferences: Uncertainty and ambiguity. *Journal of Risk and Uncertainty*, *5*, 325–370.

Chander, P., and Wilde, L. L. (1998). A general characterization of optimal income tax enforcement. *Review of Economic Studies*, *65*, 165–183.

Christiansen, V. (1980). Two comments on tax evasion. *Journal of Public Economics*, *13*, 389–393.

Cleary, D. (2011). Predictive analytics in the public sector: Using data mining to assist better target selection for audit. *Electronic Journal of e-Government*, *9*, 132–140.

Clotfelter, C. T. (1983). Tax evasion and tax rates: An analysis of individual returns. *Review of Economics and Statistics*, *65*, 363–373.

Cowell, F. A. (1981). Taxation and labour supply with risky activities. *Economica*, *48*, 365–379.

Cowell, F. A. (1985). Tax evasion with labour income. *Journal of Public Economics*, *26*, 19–34.

Cowell, F. A., and Gordon, J. P. F. (1988). Unwillingness to pay. *Journal of Public Economics*, *36*, 305–321.

Crane, S. E., and Nourzad, F. (1986). Inflation and tax evasion: An empirical analysis. *Review of Economics and Statistics*, *68*, 217–223.

Davis, J. S., Hecht, G., and Perkins, J. D. (2003). Social behaviours, enforcement, and tax compliance dynamics. *The Accounting Review*, *78*, 39–69.

De Juan, A., Lasheras, M. A., and Mayo, R. (1994). Voluntary tax compliant behavior of Spanish income tax payers. *Public Finance*, *49*, 90–115.

Eide, E. (2001). Rank dependent expected utility models of tax evasion. International Centre for Economic Research Working Paper ICER 27/2001.

Eisenhauer, J. G. (2006). Conscience as a deterrent to free riding. *International Journal of Social Economics*, *33*, 534–546.

Eisenhauer, J. G. (2008). Ethical preferences, risk aversion, and taxpayer behavior. *Journal of Socio-Economics*, *37*, 45–63.

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

Fortin, B., Lacroix, G., and Villeval, M. -C. (2007). Tax evasion and social interactions. *Journal of Public Economics*, *91*, 2089–2112.

Franklin, J. (2009). *Networks and taxpayer non-compliance.* London: H. M. Revenue and Customs.

Geeroms, H., and Wilmots, H. (1985). An empirical model of tax evasion and tax avoidance. *Public Finance*, *40*, 190–209.

Glaeser, E., Sacerdote, B., and Scheinkman, J. (1996). Crime and social interactions" *Quarterly Journal of Economics*, *111*, 507–548.

Gordon, J. P. F. (1989). Individual morality and reputation costs as deterrents to tax evasion. *European Economic Review*, *33*, 797–805.

Goyal, S. (2009). *Connections: An introduction to the economics of networks.* Princeton: Princeton University Press.

Grether, D. (1980). Bayes rules as a descriptive model: The representativeness heuristic. *Quarterly Journal of Economics*, *95*, 537–557.

Guala, F., and Mittone, L. (2005). Experiments in economics: External validity and the robustness of phenomena. *Journal of Economic Methodology*, *12*, 495–515.

Hashimzade, N., Myles, G. D., and Tran-Nam, B. (2013). Applications of behavioural economics to tax evasion, *Journal of Economic Surveys*, *27*, 941–977.

H. M. Revenue and Customs (2010). *Measuring tax gaps 2009.* London: H. M. Revenue and Customs.

Isachsen, A. J., and Strøm, S. (1980). The hidden economy: The labor market and tax evasion. *Scandinavian Journal of Economics*, *82*, 304–311.

Jackson, M. O. (2004). A survey of models of network formation: Stability and efficiency. In Demange, G. and Wooders, M. (Eds.) *Group Formation in Economics; Networks, Clubs and Coalitions* (pp. 11–57). Cambridge: Cambridge University Press.
http://www.hss.caltech.edu/~jacksonm/netsurv.pdf

James, R. (2011). Using Predictive Analytics to help target Error and Fraud in Tax Credits. *International Conference on Taxation Analysis and Research*, London, December 2011.

Kahneman, D., and Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, *47*, 263–293.

Kanbur, S. M. (1981). Risk taking and taxation: An alternative perspective. *Journal of Public Economics*, *15*, 163–184.

Kastlunger, B., Kirchler, E., Mittone, L., and Pitters, J. (2009). Sequences of audits, tax compliance, and taxpaying strategies. *Journal of Economic Psychology*, *30*, 405–418.

Kim, Y. (2003). Income distribution and equilibrium multiplicity in a stigma-based model of tax evasion. *Journal of Public Economics*, *87*, 1591–1616.

Kirchler, E., Hoelzl, E., and Wahl, I. (2008). "Enforced versus voluntary tax compliance: The "slippery slope" framework. *Journal of Economic Psychology*, *29*, 210–225.

Kolm, S.-C. (1973). A note on optimum tax evasion. *Journal of Public Economics*, *2*, 265–270.

Korobow, A., Johnson, C., and Axtell, R. (2007). An agent-based model of tax compliance with social networks. *National Tax Journal*, *60*, 589–610.

Lee, K. (2001). Tax evasion and self-insurance. *Journal of Public Economics*, *81*, 73–81.

Maciejovsky, B., Kirchler, E., and Schwarzenberger, H. (2007). Misperception of chance and loss repair: On the dynamics of tax compliance. *Journal of Economic Psychology*, *28*, 678–691.

Mittone, L. (2006). Dynamic behaviour in tax evasion: An experimental approach. *Journal of Socio-Economics*, *35*, 813–835.

Myles, G. D. (1995). *Public economics.* Cambridge: Cambridge University Press.

Myles, G. D., and Naylor, R. A. (1996). A model of tax evasion with group conformity and social customs. *European Journal of Political Economy*, *12*, 49–66.

Pestieau, P., and Possen, U. M. (1991). Tax evasion and occupational choice. *Journal of Public Economics*, *45*, 107–125.

Pyle, D. J. (1991). The economics of taxpayer compliance. *Journal of Economic Surveys*, *5*, 163–198.

Quiggin, J. (1981). Risk perception and the analysis of risk attitudes. *Australian Journal of Agricultural Economics*, *25*, 160–169.

Quiggin, J. (1982). A theory of anticipated utility. *Journal of Economic Behavior and Organization*, *3*, 323–343.

Quiggin, J., and Wakker, P. (1994). The axiomatic basis of anticipated utility: A clarification. *Journal of Economic Theory*, *64*, 486–499.

Rablen, M. D. (2010). Tax evasion and exchange equity: A reference-dependent approach. *Public Finance Review*, *38*, 282–305.

Rablen, M.D. (in press). Audit probability versus effectiveness: The Beckerian approach revisited. *Journal of Public Economic Theory*. doi: 10.1111/jpet.12062.

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

Reinganum, J., and Wilde, L. (1986). Equilibrium verification and reporting policies in a model of tax compliance. *International Economics Review*, *27*, 739–760.

Sandmo, A. (2005). The theory of tax evasion: A retrospective view. *National Tax Journal*, *58*, 643–663.

Schneider, F., and Enste, D. H. (2000). Shadow economies: Size, causes, and consequences. *Journal of Economic Literature, 38*, 77 –114.

Slemrod, J. (2007). Cheating Ourselves: The Economics of Tax Evasion. *Journal of Economic Perspectives*, 21, 25–48.

Snow, A., and Warren, R. S. Jr. (2005). Ambiguity about audit probability, tax compliance, and taxpayer welfare. *Economic Inquiry*, *43*, 865–871.

Spicer, M. W., and Lundstedt, S. B. (1976). Understanding tax evasion. *Public Finance*, *31*, 295–305.

Tesfatsion, L. (2006). Agent-based computational economics: A constructive approach to economic theory. In Tesfatsion, L. and Judd, K. L. (Eds.) *Handbook of Computational Economics* Vol. II (pp. 831–880). Amsterdam: North-Holland.

Thaler, R. H. (1994). *The winner's curse: Paradoxes and anomalies of economic life*. Princeton: Princeton University Press.

Torgler, B. (2002). Speaking to theorists and searching for facts: Tax morale and tax compliance in experiments. *Journal of Economic Surveys*, *16*, 657–683.

Trandel, G., and Snow, A. (1999). Progressive income taxation and the underground economy. *Economics Letters*, *62*, 217–222.

Traxler, C. (2010). Social norms and conditional cooperative taxpayers. *European Journal of Political Economy*, *26*, 89–103.

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

von Neumann, J., and Morgernstern, O. (1947). *Theory of games and economic behavior*. Princeton: Princeton University Press.

Wallschutzky, I. G. (1984). Possible causes of tax evasion. *Journal of Economic Psychology, 5*, 371–384.

Webley, P., Robben, H., and Morris, I. (1988). Social comparison, attitudes and tax evasion in a shop simulation. *Social Behaviour*, *3*, 219–228.

Weigel, R. H., Hessing, D. J., and Elffers, H. (1987). Tax evasion research: A critical appraisal and theoretical model. *Journal of Economic Psychology*, *8*, 215–235.

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

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

Appendix

Parameter values
Tax rate: $\tau = 0.25$
Skill spread in self-employment: $\gamma = 0.75$
Fine rate: $f = 1.5$
Weight in information exchange: $\mu = 0.75$

Probability distributions
Wage in employment: $w \sim \textit{Lognormal}\,[1.956, 0.8325]$;
$E\,[w] = 10$; $Var\,(w) = 100$
Skill in occupation 1:
$s_j^1 = \frac{1}{1-0.5\tilde{x}}$, $\tilde{x} = U(0,1)$
Income in occupation 1:
$s_j^1 y^1$, $y^1 \sim \textit{Lognormal}\,[1.3785, 1.1840]$;
Skill in occupation 2:
$s_j^2 = \frac{1}{1-0.5\tilde{x}}$, $\tilde{x} = U(0,1)$
Income in occupation 2:
$s_j^2 y^2$, $y^2 \sim \textit{Lognormal}\,[1.0430, 1.4813]$;
Risk aversion: $\rho \sim U\,[0.1, 5.1]$
Initial belief on audit probability: $p_0 \sim U\,[0,1]$
Importance assigned initially to social custom: $\chi_0 \sim U\,[0,1]$
Value of social custom: $z \sim U\left[0, 3 \times 10^{-5}\right]$

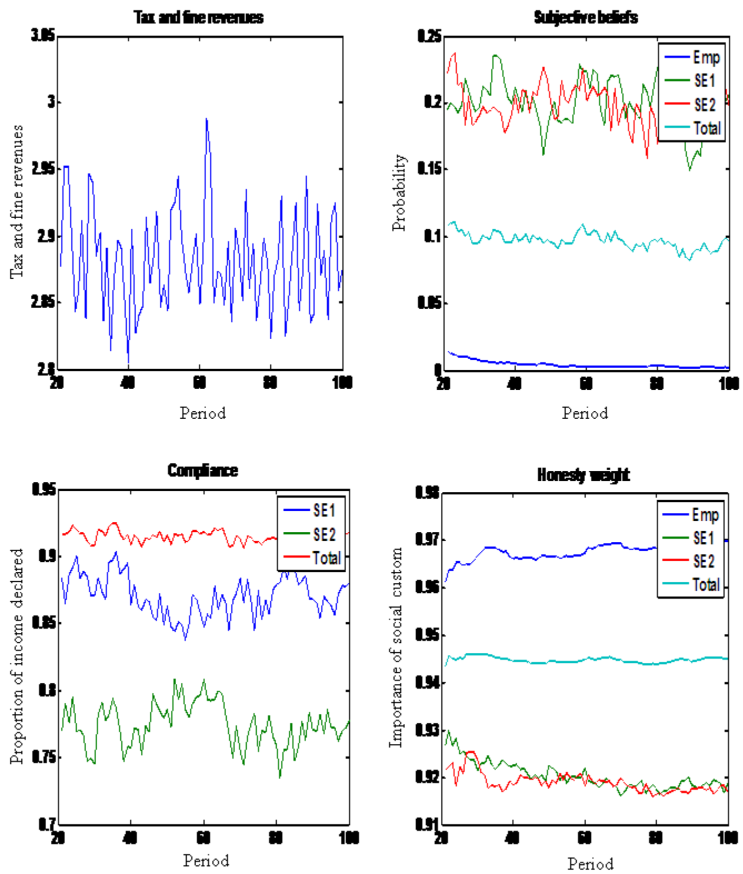

Figure 1: Baseline with Random Audits

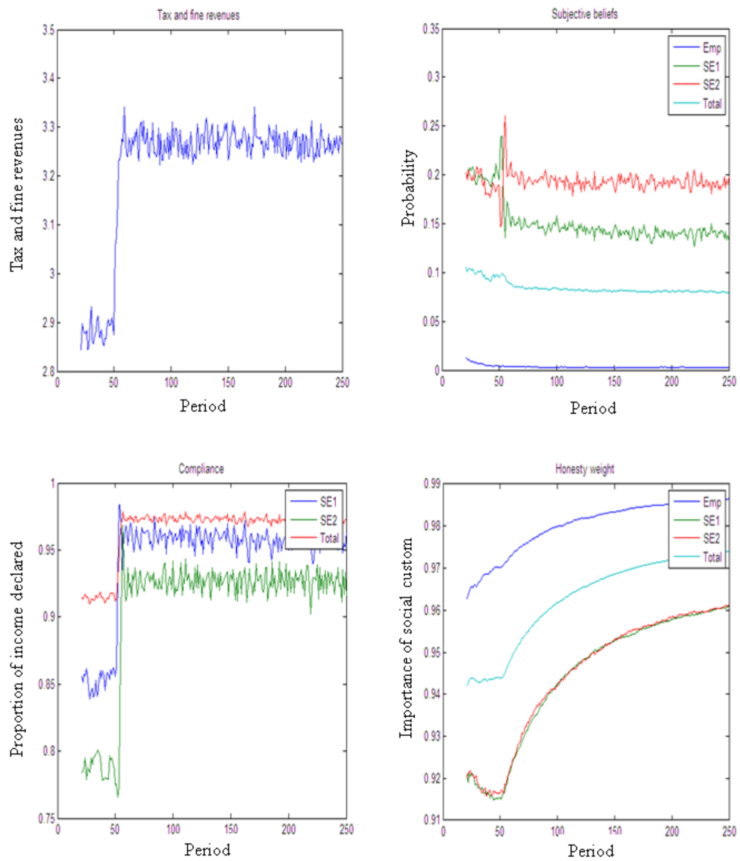

Figure 2: Largest Evaders Targeted Audits

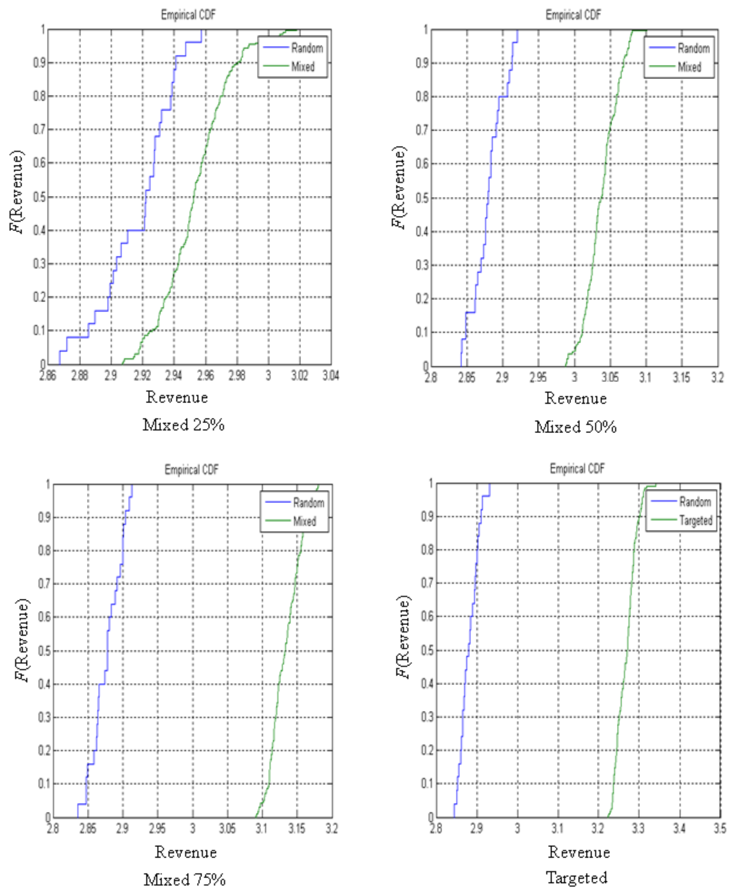

Figure 3: Largest Evaders Empirical CDFs

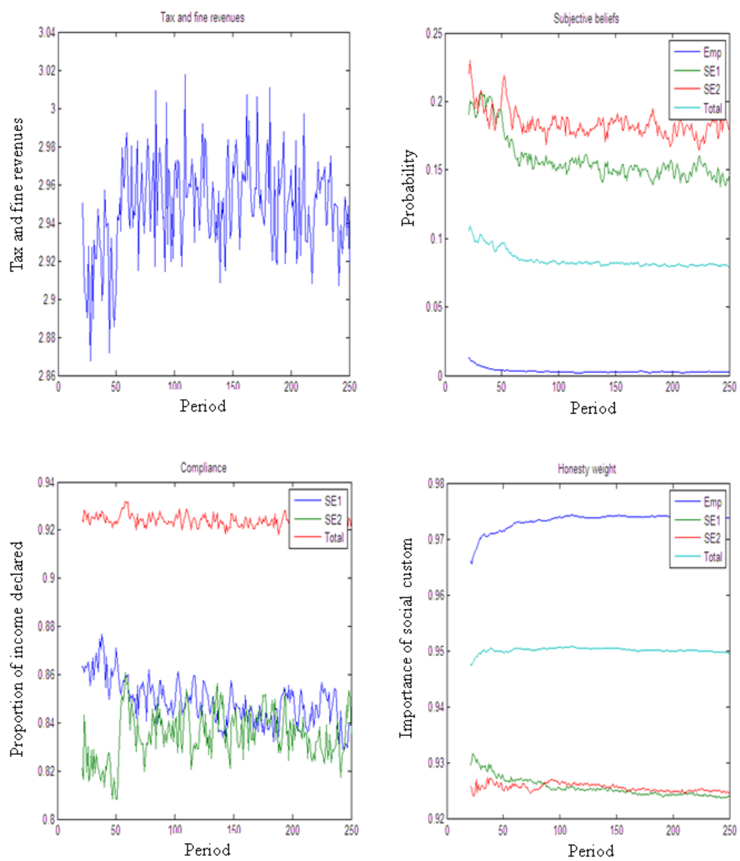

Figure 4: Largest Evaders 25% Mixed

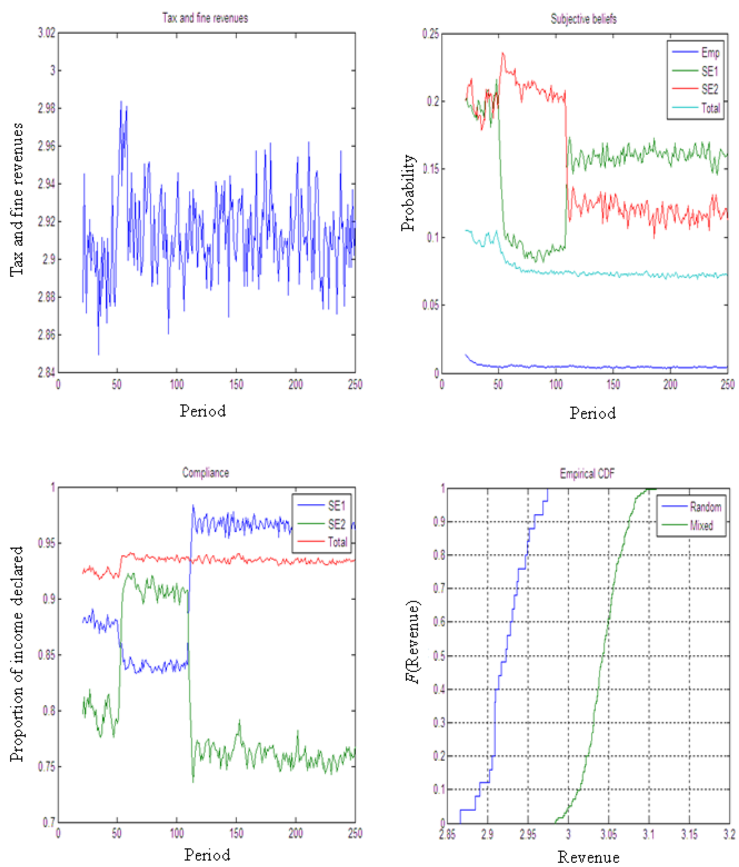

Figure 5: Most likely Evaders 50% Mixed