Predictive Analytics and the Targeting of Audits

Nigar Hashimzade (Durham University) Gareth Myles (University of Exeter)

October 2014

Introduction

- The role of *predictive analytics* is to identify the best targets for audit
- Predictive analytics are used by the IRS, HMRC etc.
- Various methods are used including credit scoring and econometric analysis
- We want to explore the effects of predictive analytics and whether they can improve on the other audit strategies
- The analysis compares the outcome of predictive analytics based on tax return data with that of random audits
- This is undertaken within an agent-based simulation

Attitudes, Beliefs, and Opportunities

- The taxpayers in the model choose occupation, degree of compliance, and interact within a social network
- The compliance decision is based on *attitudes*, *beliefs*, and *opportunities* and recognizes the *social setting* for the decision
- A central component of the model is the endogenous formation of *attitudes* and *beliefs*
- These are determined by experience and through information exchange within the *social network*
- The modelling of these features is now described

Individual Characteristics

- 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 parameters

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

- The parameter are independently distributed across taxpayers
- The first five parameters are assigned to the taxpayer at the outset of the analysis and remain constant
- The remaining four parameters are updated through interaction with the revenue service and with other taxpayers in the social network

Occupational Choice

- The model allows each individual to make a choice of occupation (a generalization of Pestieau and Possen, 1991)
- *Employment* is safe (wage is fixed) but tax cannot be evaded (withholding, third-party reporting)
- *Self-employment* is risky but provides an opportunity to evade
- In each period, t, every individual chooses their preferred occupation and, once income is known, the optimal level of evasion
- If employment is chosen the wage, w_j, is obtained with certainty

Occupational Choice

- The outcome of self-employment for individual j in occupation α at time t is given by $s_j^{\alpha} y_{j,t}^{\alpha}$
- $y_{j,t}^{\alpha}$ is a random variable described by the probability distribution function $F^{\alpha}(\cdot)$
- It is assumed that $\mu(y_1) < \mu(y_2)$ and $\sigma^2(y_1) < \sigma^2(y_2)$
- The evasion level is chosen after income from self-employment is realized

Occupational Choice

- The expected payoff from the optimal evasion in occupation α given a realization $y_{j,t}^{\alpha}$ is

$$V_e^{\alpha}\left(y_{j,t}^{\alpha}\right) = \max_{E_j^i \in \left[0, s_j^i y_{j,t}^i\right]} \left\{ p_{j,t}^{\alpha} U\left(s_j^{\alpha} y_{j,t}^{\alpha} - f\tau E_j^{\alpha}\right) \right.$$

$$\left. + (1 - p_{j,t}^{\alpha}) U\left(s_j^{\alpha} y_{j,t}^{\alpha} + \tau E_j^{\alpha}\right) + \chi_{j,t} z_j \mathbf{1}_{\left[E_j^i = 0\right]} \right\}$$

- $f > 1$ is the fine and $\mathbf{1}_{[A]}$ an indicator function with value one if A is true, zero otherwise
- The payoff from the social custom is obtained only if tax is paid in full
- The expected payoff from the compound lottery describing occupation α is then

$$V^{\alpha} = \int_{Y^{\alpha}} V_e^{\alpha}(y)\, dF^{\alpha}(y).$$

- The choice of occupation is made by selecting the maximum of $\left\{V^0, V^1, V^2\right\}$.

Attitudes and Beliefs

- The analysis of tax evasion has demonstrated two important features:
  - The social setting influences the evasion decision (*attitudes*)
  - The probability of audit is subjective not objective (*beliefs*)
- We have incorporated these into the simulation by adding learning within a *social network*
- Individuals meet with their contacts in the network and meetings allow exchange of information on beliefs
- This can explain why social groups have different behavior with respect to tax evasion

Network and Meetings

- The network is described by a symmetric matrix A of $0s$ and $1s$ (bi-directional links)
- In each period a random selection of meetings occur described by a matrix C of zeros and ones
- Individuals i and j meet during a period if $A_{ij}C_{ij} = 1$
- 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 $\alpha, \beta = 0, 1, 2$

Subjective Beliefs

- Beliefs are updated in two ways
- Taxpayer j in occupation α in period t adjusts belief $p^{\alpha}_{j,t}$ if audited or if not audited
- The taxpayer may meet with a contact in the network
- If the meeting is with a taxpayer in occupation β information is exchanged with probability $q_{\alpha\beta}$
- This information is then used to update the belief $p^{\beta}_{j,t}$

Subjective Beliefs

- The *target effect* assumes that taxpayers feel marked as targets if they are audited
- Those not audited in a period believe they are less likely to be audited in the next period
- If audited in period t, an individual's belief about being audited in the next period is raised to the level P, otherwise it decays
- The updating rule for the subjective probability is

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

  where $A_{j,t} = 1$ if taxpayer j was audited in period t and $A_{j,t} = 0$ otherwise
- We refer to the case of $P = 1$ as the maximal target effect

Subjective Beliefs

- The *bomb-crater* effect assumes taxpayers audited in one period believes they are unlikely to be audited in the next
- If not audited the belief rises over time
- The process is described by

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

- We refer to $P = 0$ as the maximal bomb-crater effect

Subjective Beliefs

- 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

$$\begin{aligned} p_{j,t+1}^{\beta} &= \mu \tilde{p}_{j,t}^{\beta} + (1-\mu)\, \tilde{p}_{i,t}^{\beta} \\ p_{j,t+1}^{\gamma} &= \tilde{p}_{j,t}^{\gamma}, \quad \gamma \neq \beta \end{aligned}$$

Attitudes

- The importance of the social custom is determined by interaction in the social network
- Each individual is randomly assigned a level of importance, χ_0^i, at time 0
- This value is then updated each period *if* there is an information exchange between two individuals
- The updating process is described by

$$\chi_{t+1}^i = \frac{1}{X(i)+1}\left[\chi_t^i X(i) + \mathbf{1}_{\left[E_t^j=0\right]}\right]$$

  where $X(i)$ is the number of previous meetings for i at which information was exchanged
- $\chi_{t+1}^i > \chi_t^i$ if information is exchanged with an honest taxpayer and $\chi_{t+1}^i < \chi_t^i$ if information is exchanged with an evader

Simulation Process

- Individual characteristics are randomly drawn at the outset
- The simulation then iterates the following steps:
  - Occupation is chosen
  - Incomes are realized (as random draws in self-employment) and the evasion decision is made
  - The tax authority audits and punishes any evasion that is detected
- For each iteration the outcome with honesty and with evasion are calculated
- 4000 individuals in the simulation, 50 periods random, 200 predictive, 200 mixed random and predictive

Tobit

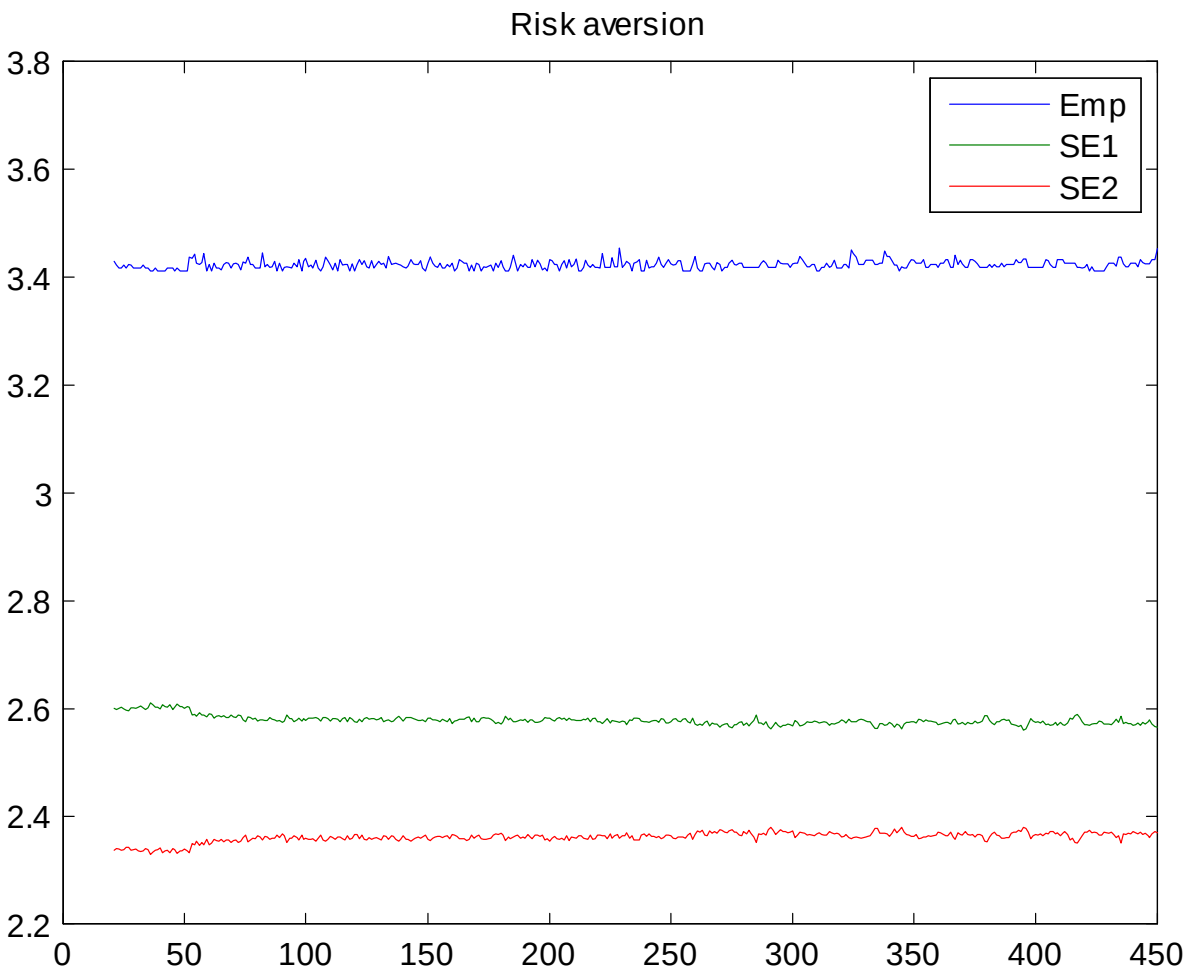
- The simulation uses random audits for the first 50 periods
- The audit data from periods 46-50 is collected and used to run a Tobit (censored) regression
- The amount of non-compliance is regressed on *occupation*, *declaration*, and *audit history*
- The estimated equation is used to predict non-compliance
- For periods 51-250 the top 5 percent are audited and audit outcomes used to update regression
- For periods 251-450 the top 2.5 percent are audited and 2.5 percent are randomly audited

Tobit

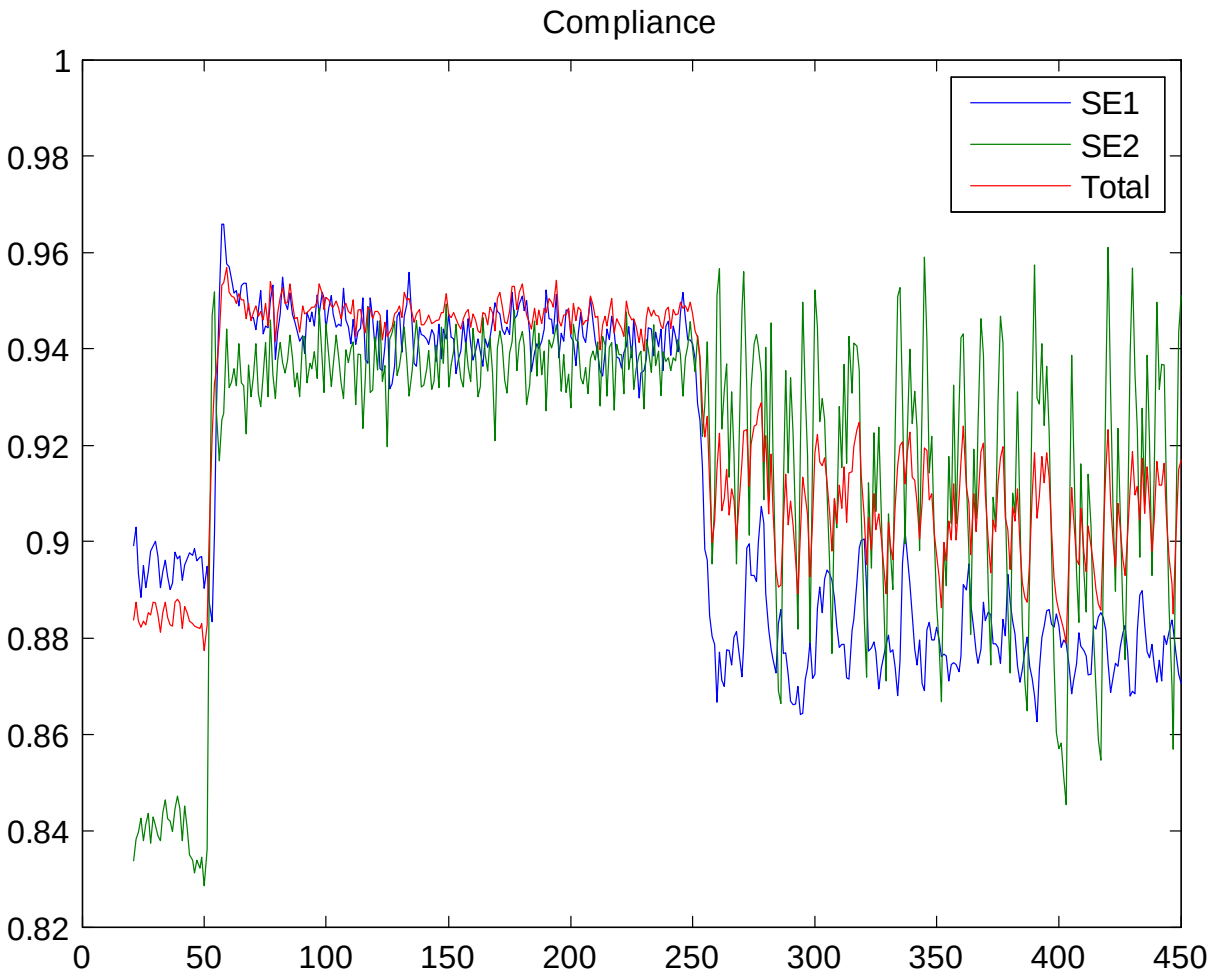
Variable	**ME (individual)**
Intercept	17.2871
Declared Income	-4.1267
Previous audit	-5.5072
Self-employment 1	-8.1844

Marginal effects of explanatory variables upon under-reported income

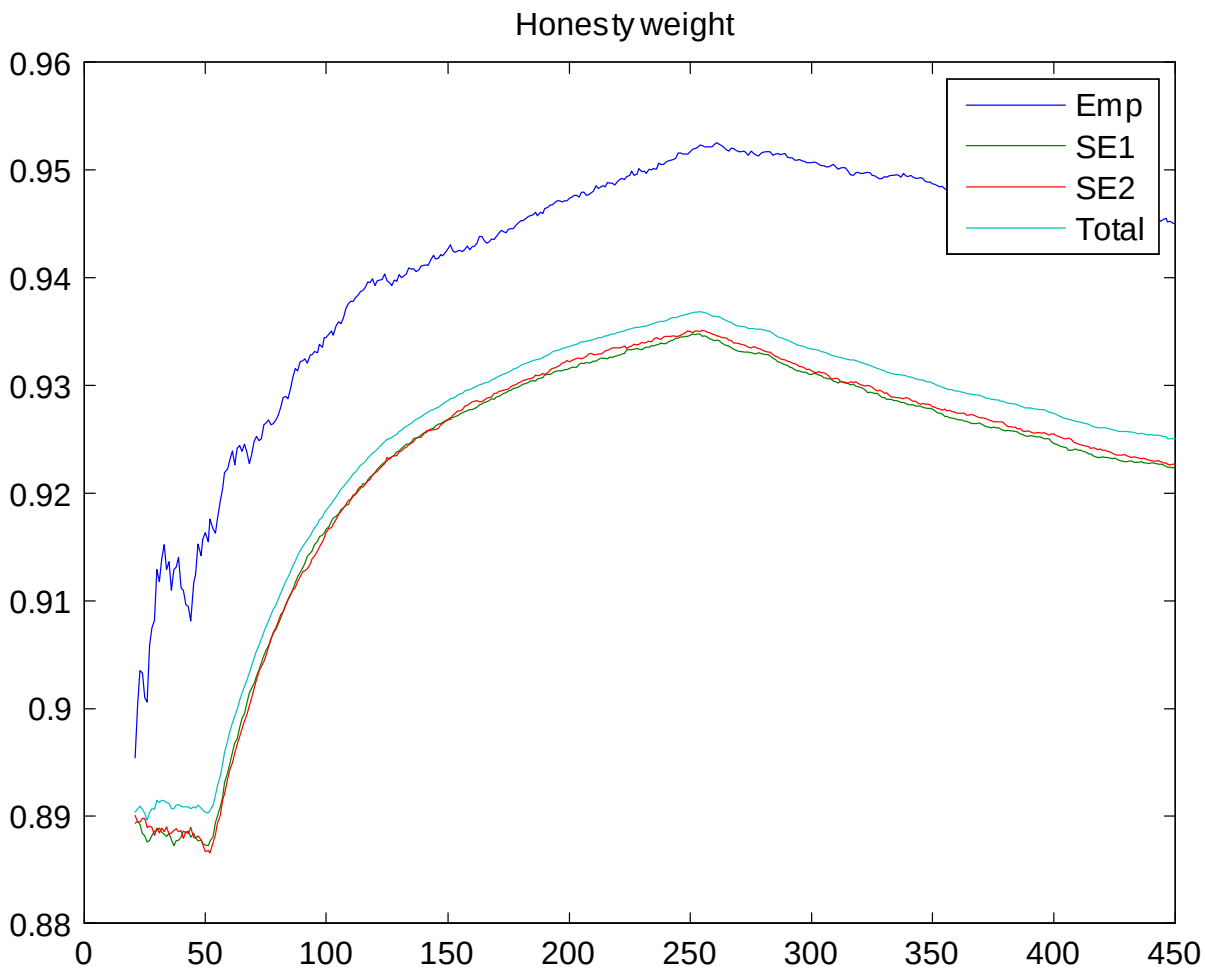
Tobit


Risk aversion
Emp
SE1
SE2
3.8
3.6
3.4
3.2
3
2.8
2.6
2.4
2.2
0
50
100
150
200
250
300
350
400
450

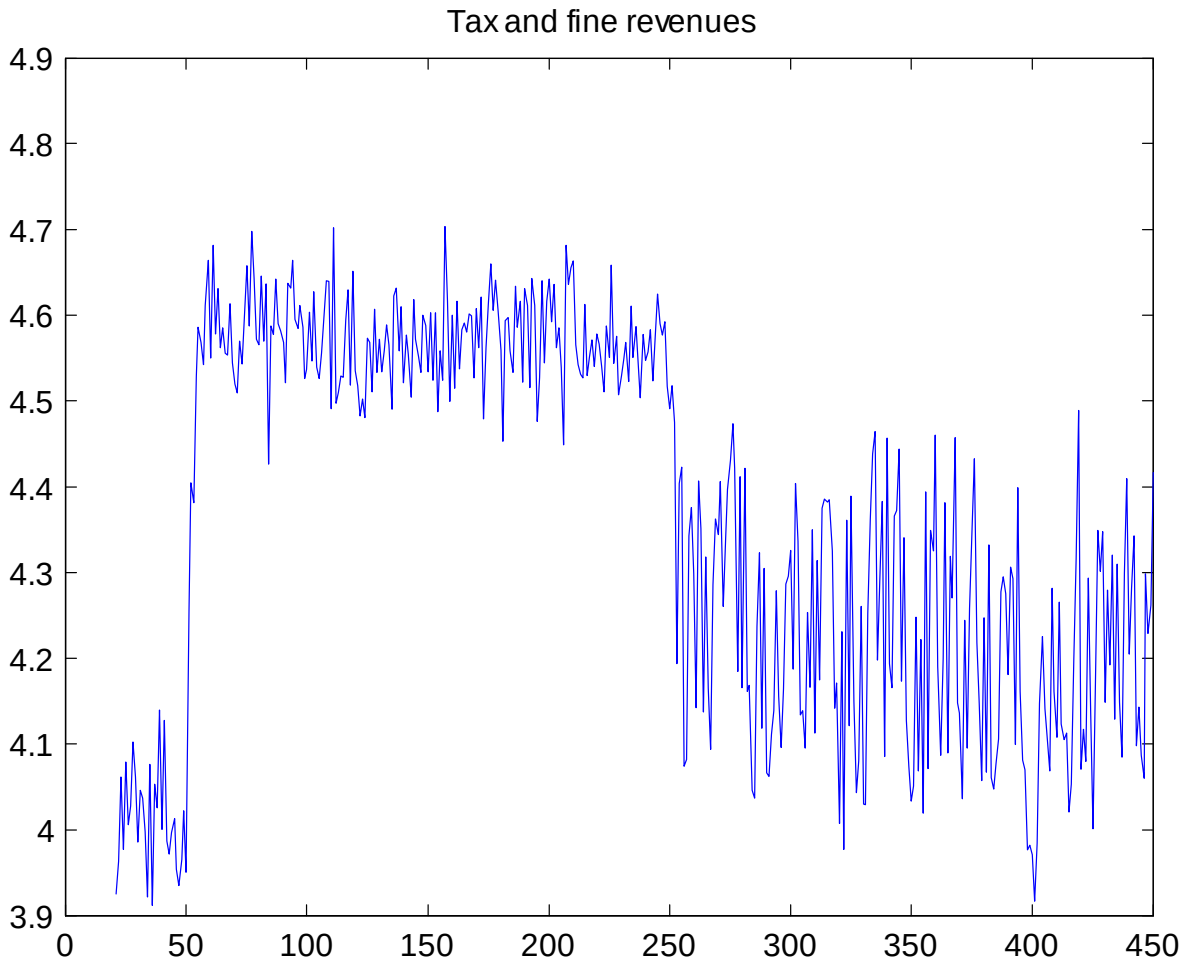
Tobit


Compliance
SE1
SE2
Total
1
0.98
0.96
0.94
0.92
0.9
0.88
0.86
0.84
0.82
0
50
100
150
200
250
300
350
400
450

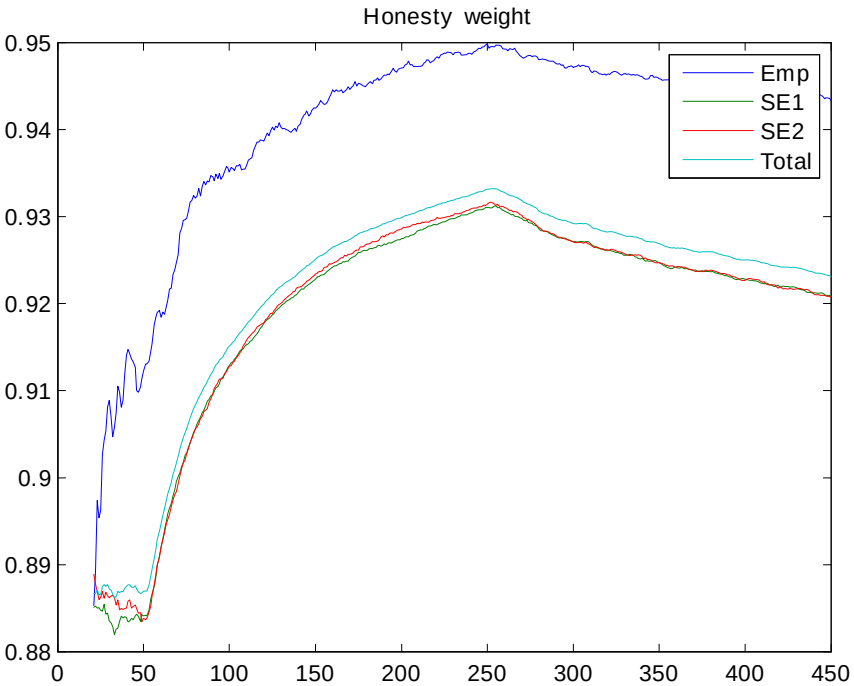
Tobit

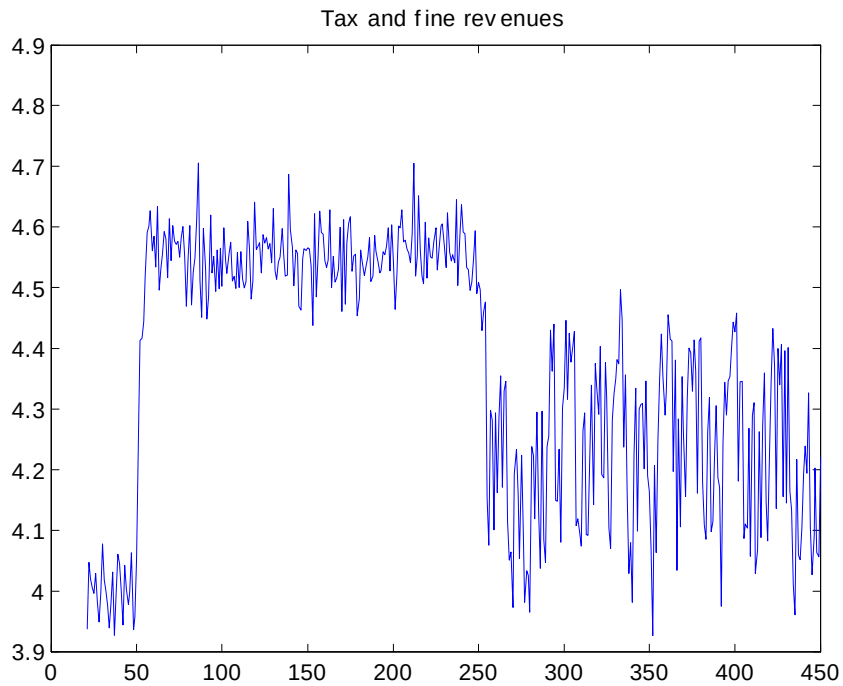

Honesty weight
0.96
0.95
0.94
0.93
0.92
0.91
0.9
0.89
0.88
0
50
100
150
200
250
300
350
400
450
Emp
SE1
SE2
Total

Tobit

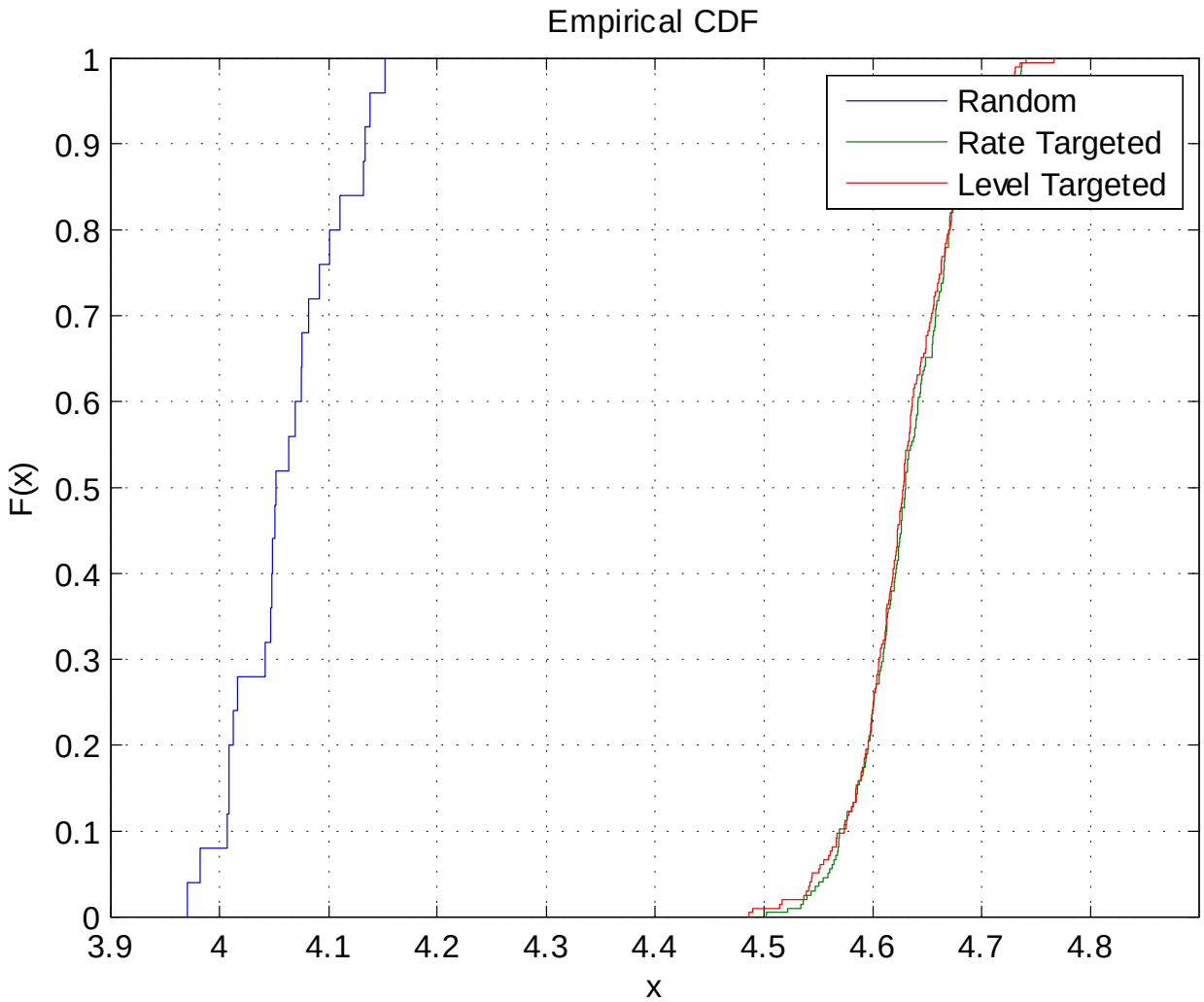

Tax and fine revenues
4.9
4.8
4.7
4.6
4.5
4.4
4.3
4.2
4.1
4
3.9
0
50
100
150
200
250
300
350
400
450

Logit


Honesty weight
Emp
SE1
SE2
Total
0.95
0.94
0.93
0.92
0.91
0.9
0.89
0.88
0
50
100
150
200
250
300
350
400
450


Tax and fine revenues
4.9
4.8
4.7
4.6
4.5
4.4
4.3
4.2
4.1
4
3.9
0
50
100
150
200
250
300
350
400
450

Logit, Tobit, and Random



Age and Risk Aversion

Extended model

- There is now common agreement in the literature that individual risk aversion does not remain constant over the life span
  - Young people are on average less risk averse than old people
  - Risk aversion increases with age
- A revenue service can use age as one of the risk factors in targeted audits
  - Age can serve as a proxy of the willingness to take risk
  - (Can use number of years of filing)
- We now extend the model by assuming the CRRA coefficient increases with age

Age and Risk Aversion

Assumptions

- As a benchmark, assume ρ is determined by a linear function with an additive stochastic component

$$\rho_{it} = a + b\frac{Age_{it}}{MaxAge} + \varepsilon_{it}, \;\; E\left[\varepsilon_{it}\right] = 0$$

  - A linear-quadratic form (concave or convex) gives qualitatively similar results

- Agent i works in period t if $Age_{it} \leq MaxAge$ and retires otherwise
  - In the simulations we also randomize the retirement age

$$MaxAge_{it} = MaxAge \times (1 + u_{it}), \quad E\left[u_{it}\right] = 0$$

- Retired agents are replaced with agents of age zero who have new individual characteristics
  - An alternative version also includes inheritance of characteristics

Age and Risk Aversion

Audit strategies

- The revenue service observes (1) declared income, (2) previous audit history (caught evading in previous period or not), (3) occupation and (4) age
  - The first three variables were used in the previous simulations
- Age is used in addition to or instead of occupation
  - With many occupations the number of regressors can be large
  - Relative riskiness may be unknown or not differ much
- As before, we compare the effectiveness of random, targeted and mixed audits

Age and Risk Aversion

Simulation results

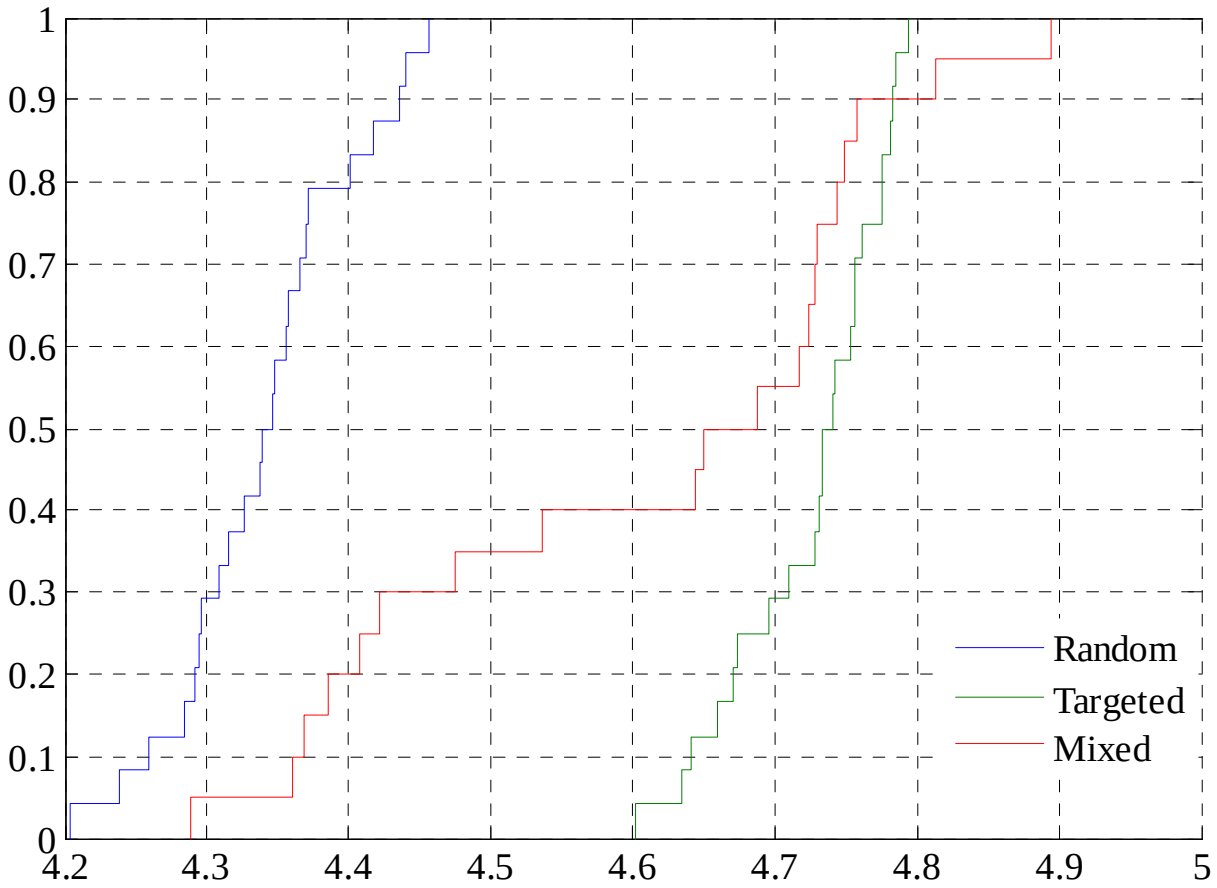
- Parameters:

$$\begin{aligned} a &= 0.1;\ b = 5;\ \varepsilon_{it} \sim 0.5 \times U\left[-0.5, 0.5\right];\ MaxAge = 25; \\ u_{it} &\sim 0.5 \times U\left[-0.5, 0.5\right];\ MaxAge_{it} \in \left[20, 29\right];\ Age_{i1} \sim UI\left[0, 25\right] \end{aligned}$$

- Three specifications: (1) occupation; (2) age; (3) occupation and age
- In (1) and (3) random audits are the least effective
- In (1) and (2) targeted audits are the most effective except for the upper tail
- In (3) targeted audits are the most effective

Age and Risk Aversion

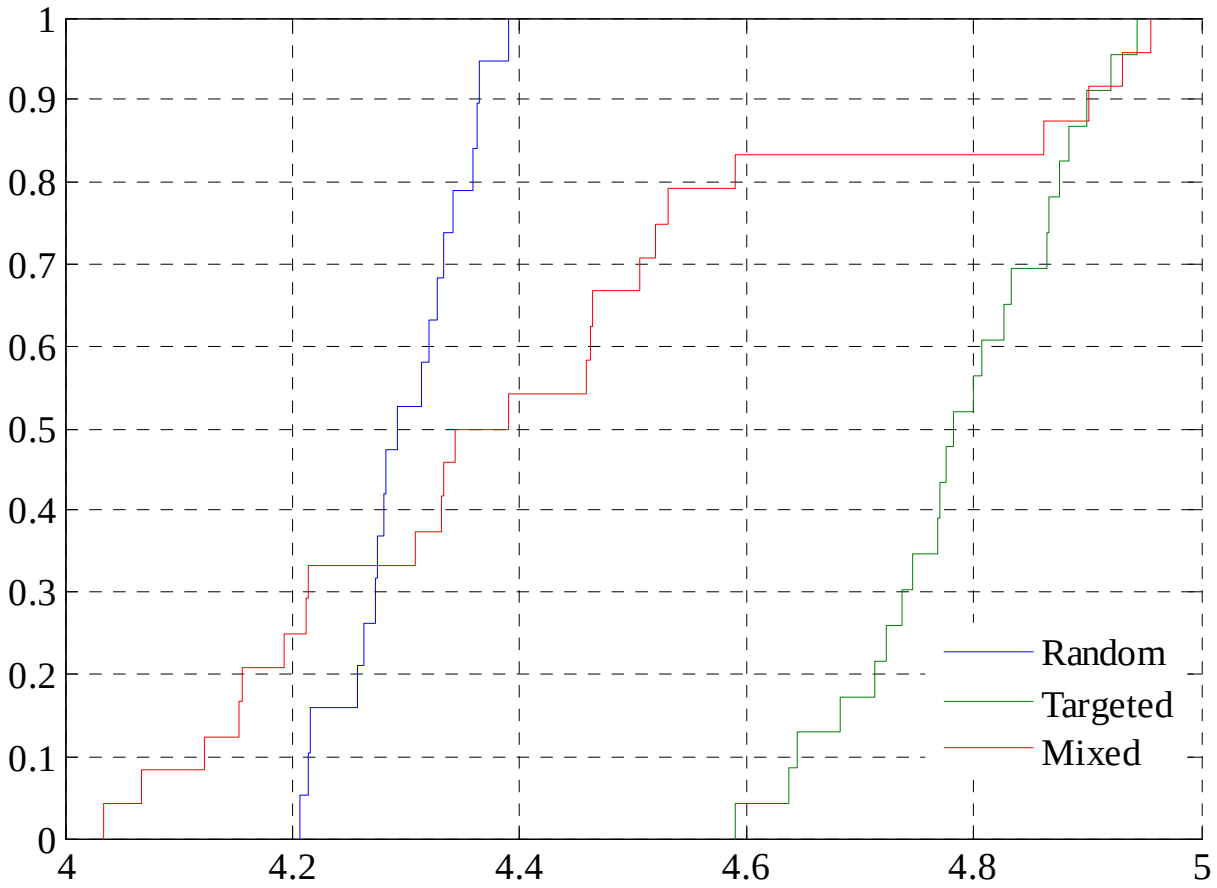
Revenue distribution



Risk factors: declaration, past audit, occupation.

Age and Risk Aversion

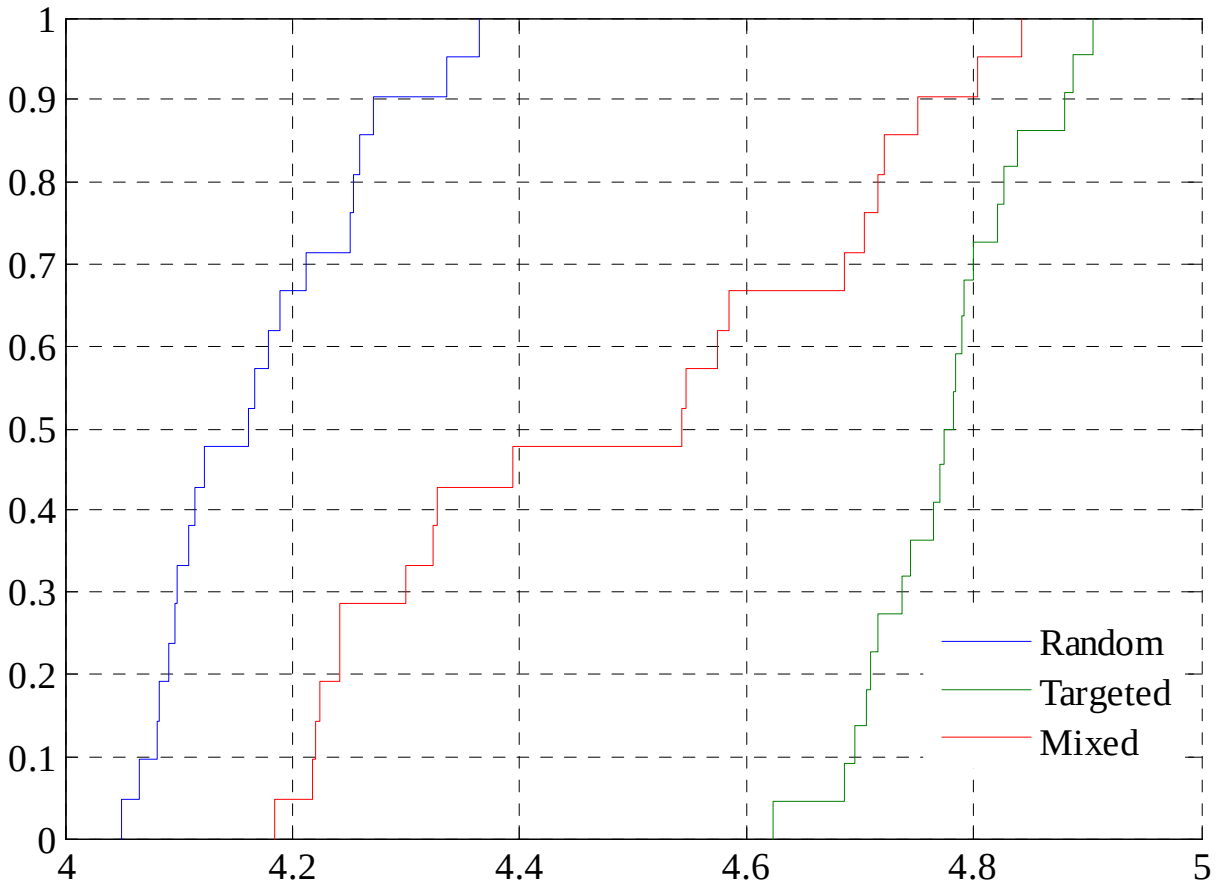
Revenue distribution



Risk factors: declaration, past audit, age.

Age and Risk Aversion

Revenue distribution



Risk factors: declaration, past audit, occupation, age.

Predictive Analytics

- The results show clearly that the use of predictive analytics increase tax and fine revenue
- Underlying this is an increase in the honesty weight when the predictive analytics operate
- Compliance is not uniformly increased in occupational groups if there is some randomness
- Extending to the dual probabilities does not affect the conclusion

Conclusions

- Agent-based modelling is a useful tool for testing policies
- The modelling can incorporate recent advances in the theory of compliance
- Our work emphasizes the role of attitudes, beliefs, and opportunities
- Compliance behavior can vary significantly across occupational groups
- Predictive analytics is successful in encouraging compliance and increasing revenue