Targetting Audits Using Predictive Analytics

Gareth Myles with Nigar Hashimzade, Frank Page, and Matt Rablen

Exeter and IFS

July 2013

Introduction

- The economic analysis of tax compliance has two objectives:
  - To explain and predict behaviour
  - To design beneficial interventions
- Different methodologies can contribute to this objective:
  - Theory
  - Empirical analysis
  - Experimentation
- This talk will focus on an additional methodology: *agent-based modelling*

Introduction

- The talk begins with an introduction to agent-based modelling
- Successful application of agent-based modelling requires a credible model of individual compliance
- So the literature on modelling compliance is reviewed
- The talk is completed by reviewing my work with Nigar Hashimzade, Frank Page, and Matt Rablen
  - *Auditing rules*
  - *Predictive analytics*

Agent-Based Modelling

- An agent-based model:
  - Creates a set of agents
  - Assigns abilities, objectives, and knowledge
  - Allows them to interact
  - Observes the outcome
- The creation and interaction takes place in a computer simulation
- Parameters can be varied to test the effect on the outcome
- Such models can describe natural situations or economic situations

Sheep and Wolves

- A famous agent-based model of nature is that of sheep and wolves
- Wolves and sheep wander randomly around the landscape
- The wolves look for sheep to prey on
  - Each step costs wolves energy so they must eat sheep
  - When they run out of energy they die
- Sheep eat grass and reproduce
- The analysis simulates the evolution of the populations

Allingham-Sandmo[1]

- The same software can support a basic tax evasion model:
  - Apply the Allingham-Sandmo model of evasion choice
  - Adopt a random audit strategy
  - Track the degree of compliance
- Policy experiments permit the effect of interventions to be judged
- Provides a starting point for more detailed analysis

[1] Thanks are due to Kim Bloomquist, IRS, for allowing use of the programme

Limitations

- There are several limitations:
  - Most free software does not permit complex optimization
  - The implications of the preferences do not fit the facts
  - Interventions can be more sophisticated than random audits
- The first is solved through the use of alternative software (Matlab)
- The second motivates a review of the recent literature on compliance
- The analysis of sophisticated interventions is our current research question

Focussing on Choices

- Our research focuses on modelling the *choice behaviour* behind the compliance decision
- We aim to integrate the best of current theory to match evidence
- The intention is to permit the exploration of policy interventions
- The models can use artificial data or be calibrated to actual data
- The next sections develop the components of the model

Individual Compliance Behaviour

- Research on compliance behaviour has built on the basic model of Allingham-Sandmo (1972)
- The evasion level is chosen to maximize expected utility

$$\mathcal{E}U = pU(Y[1-t] - tfE) + [1-p]U(Y[1-t] + tE)$$

- Where:
  - p is the probability of audit
  - Y is income
  - t is the tax rate
  - f is the fine levied on tax evaded
  - E is the amount of evasion

Limitations

- There are three basic problems with the predictions of this model
- *First*, $E > 0$ if $p < \frac{1}{1+F}$ which is satisfied for practical values (f is at most 2, so $E > 0$ if $p < 1/3$)
- *Second*, decreasing absolute risk aversion is sufficient for $\frac{dE}{dt} < 0$
- *Third*, the source of income also determines the *opportunity* for evasion:
  - Third-party reporting
  - Withholding
- Solutions proposed to improve the predictions include appeal to *non-expected utility theory* and to *social customs*

Non-Expected Utility

- A general form of non-EU choice theory is

$$V = w_1(p, 1-p)v(Y[1-t] - tfE) + w_2(p, 1-p)v(Y[1-t] + tE)$$

- Several alternatives have been proposed:
  - *Rank Dependent Expected Utility* imposes structure on the weighting functions
  - *Prospect Theory* uses weights, changes payoff functions, and comparison to a reference point
  - *Non-Additive Probabilities* do not require the normal consistency of aggregation for probabilities
  - *Ambiguity* permits uncertainty over the probability of outcomes
- The weighting functions (or *beliefs*) can improve predictions but still do not give $\frac{dE}{dt} > 0$
- And these alternatives have their own shortcomings (Hashimzade, Myles and Tran-Nam, 2012)

Prospect Theory

- Yaniv (1999), al Nowaihi and Dhami (2001), and Bernasconi and Zanardi (2004) use variants of prospect theory
- Consider the standard Kahneman-Tversky value function

$$v(z) = \begin{cases} z^{\beta}, & \text{if } z > 0 \\ -\gamma\left(-z^{\beta}\right), & \gamma > 1, \text{ if } z < 0 \end{cases}$$

- And choose the reference point as the correct tax payment, $Y\left[1-t\right]$
- The payoff function becomes

$$V = E^{\beta} t^{\beta} \left[w_2 - w_1 \gamma f^{\beta}\right]$$

- So evasion is all or nothing (a consequence of the non-concave objective)

Social Customs

- A social custom is an informal rule of behaviour that summarizes the *attitude* toward compliance
- A loss of utility is incurred if the custom is broken

$$V = \begin{cases} U(Y[1-t]) + \chi^i, & \text{if } E = 0 \\ \mathcal{E}U, & \text{if } E > 0 \end{cases}$$

- There will be a cutoff χ^* such that $\chi^i < \chi^* \Longrightarrow E > 0$ and $\chi^i > \chi^* \Longrightarrow E = 0$
- If $\chi^i = \chi^i(m, E)$, (m the proportion of population evading) evasion becomes a social decision
- Myles and Naylor (1996) show that $\chi^i_m(m, E) < 0$ opens the possibility of multiple equilibria
- For some specifications it is also possible for $\frac{dE}{dt} > 0$

Attitudes, Beliefs, and Opportunities

- From these observations:
  - We do not need to feel bound by expected utility using objective probability
  - Similarly, there is no need to be restricted by any of the particular alternatives
- In short, we happily mix subjective beliefs with convenient functional forms
- Our model of the compliance decision combines *attitudes*, *beliefs*, and *opportunities* and recognizes the *social setting* for the decision
- Three uses of agent-based models are now described:
  - The effect of *opportunities* are considered within the Allingham-Sandmo framework
  - Next the endogenous development of *attitudes* and *beliefs* within a *social network* is added
  - The final step is to review the effect of *predictive analytics* on audit outcomes

Opportunities

- 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
- An individual is described by $\{w, \rho, s_1, s_2\}$:
  - $w =$ wage in employment
  - $\rho =$ (relative) risk aversion
  - $s_i =$ skill in self-employment occupation i

Occupational Choice

- The outcome of self-employment is $s_i y_i$ where y_i is drawn from a lognormal distribution
- 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 known
- With outcome $Y_i = s_i y_i$ the amount evaded E_i is determined by

$$\max \mathcal{E} U_i = pU([1-t]Y_i - ftE_i) + (1-p)U([1-t]Y_i + tE_i)$$

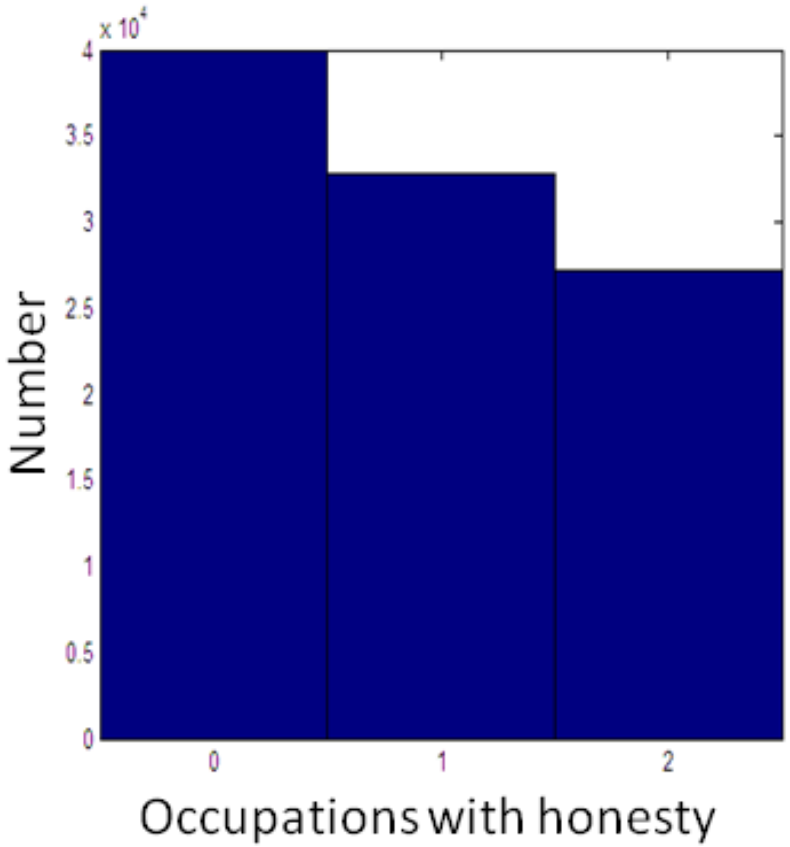
- The occupation offering highest (expected) utility is chosen

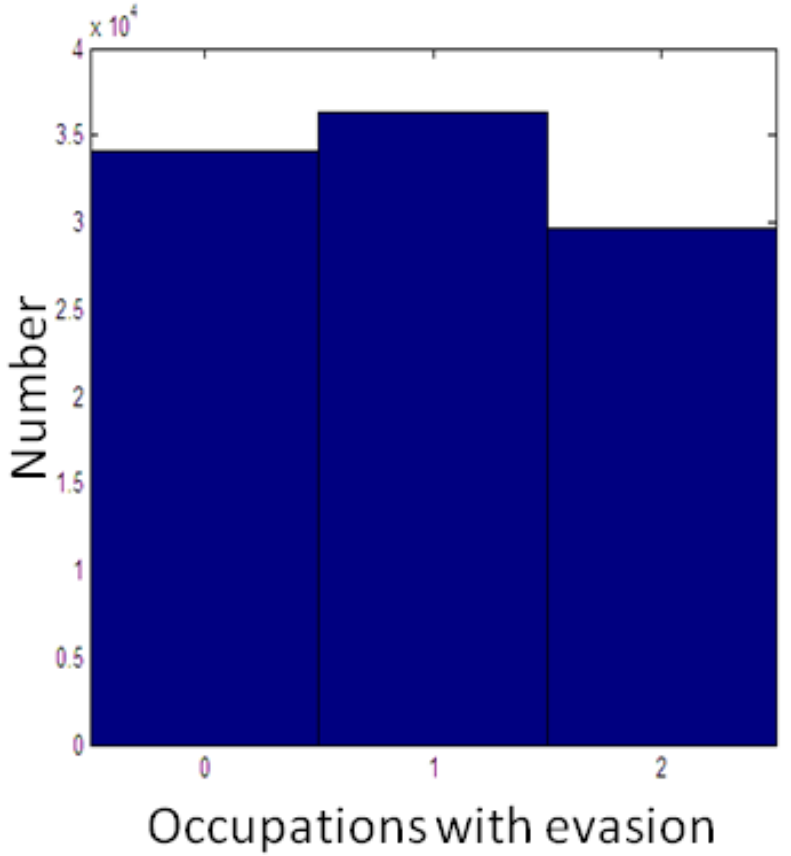
Simulation Process

- Individual characteristics $\{w, \rho, s_1, s_2\}$ 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
- 1000 individuals in the simulation, 100 iterations and data averaged across iterations

Evasion and Risk-Taking

- The distribution of occupational choices shifts away from the safe occupation
- There is more occupational risk-taking when evasion is possible

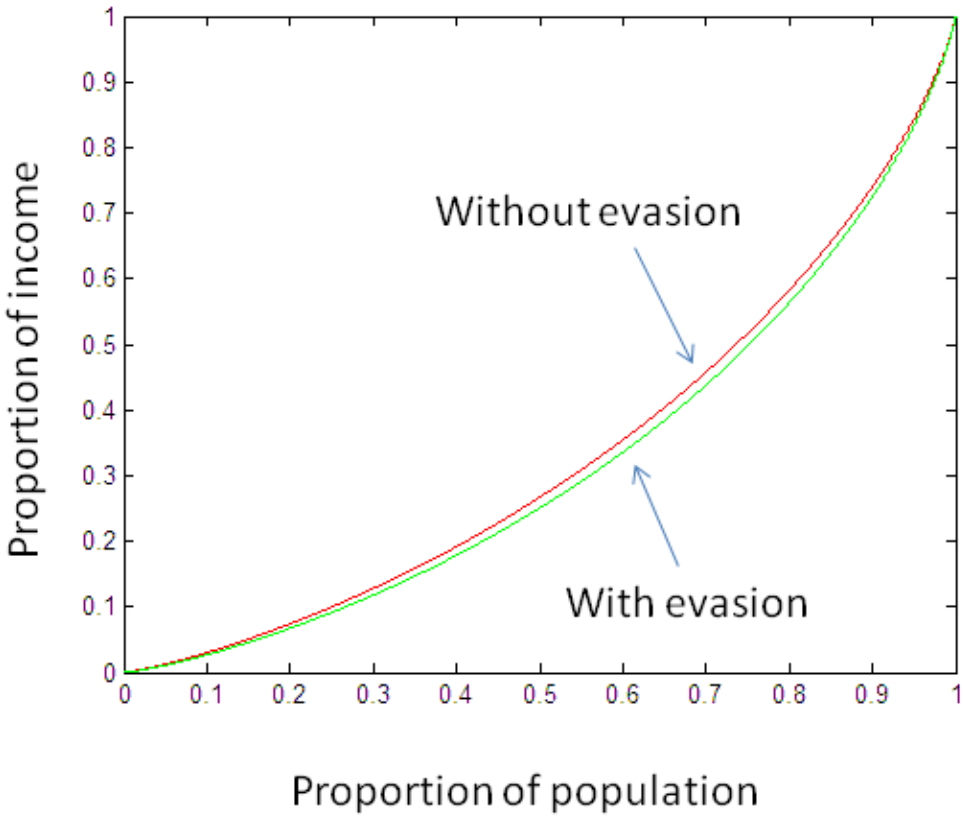




Evasion and Income Distribution

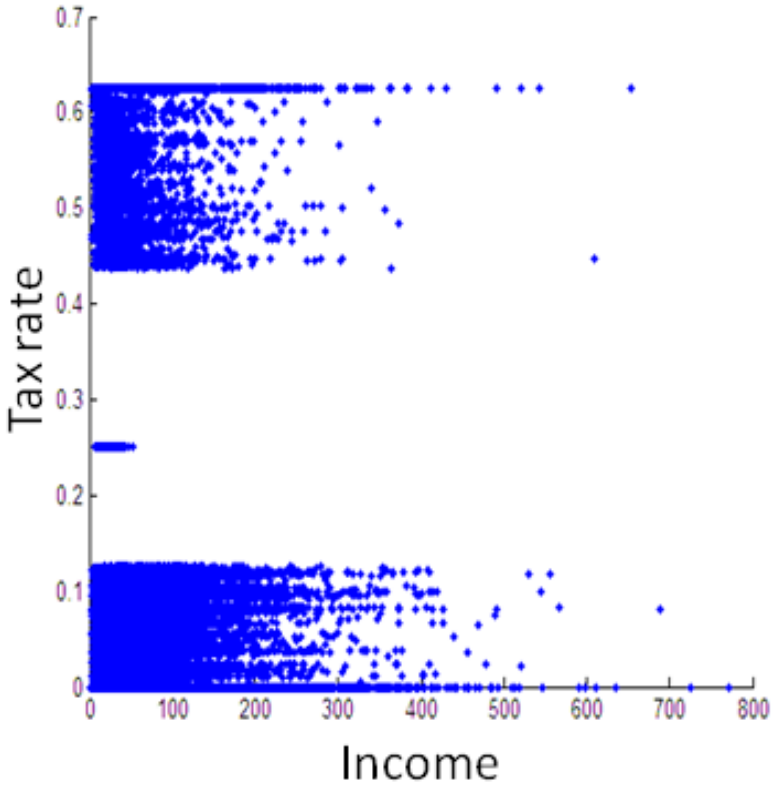
- Evasion increases mean income (after taxes and fines) and the inequality of income

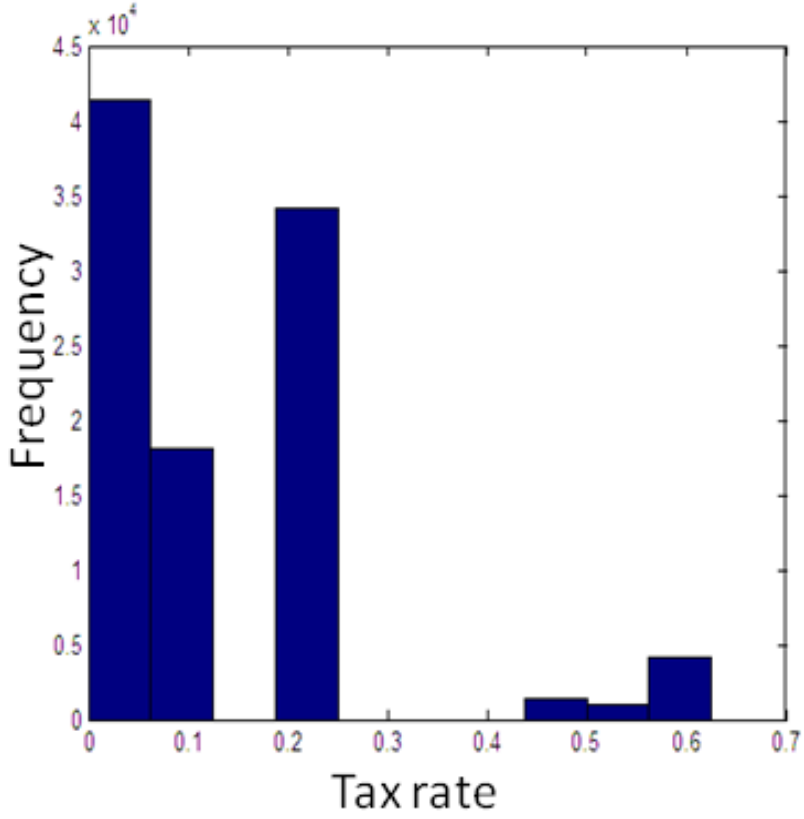
	Honesty	Evasion
Mean Income	27.72	32.77
Gini Coefficient	0.464	0.492



Evasion and the Effective Tax Rate

- The flat tax of 0.25 is undermined by evasion and punishments
- The distribution of effective tax rates is unrelated to income





Effective tax rate

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 behaviour 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 $A_{ij}C_{ij} = 1$
- At a meeting of i and j there is a probability that information is exchanged
- The probability of information exchange depends on the occupational groups to which i and j belong
- The probabilities are given by p_{ij} where $i, j = e, 1, 2$, and $p_{ii} > p_{ij}$, all i, j, $i \neq j$

Audits and Belief Updating

- The belief about the probability of audit is determined by audits and interaction
- Occupational choice in period t is made on the basis of the belief p_t^i
- The updating effect of an audit is

$$\tilde{p}_t^i = X_t^i P + \left(1 - X_t^i\right) d\left(p_t^i\right), \ P \in [0,1]$$

  where $X_t^i = 1$ if i was audited in t and $X_t^i = 0$ otherwise
- Two different processes for the formation of subjective beliefs:
  - *Target effect*: $P = 1$ and $d\left(p_t^i\right) = \delta p_t^i$, $\delta \in [0,1]$ (rise, then decay)
  - *Bomb-crater effect* (Guala and Mittone, 2005): $P = 0$ and $d\left(p_t^i\right) = p_t^i + \delta\left(1 - p_t^i\right)$, $\delta \in [0,1]$ (fall, then rise)
- The evidence on which is correct is not compelling

Information Exchange

- Individuals meet *after* audits take place
- *If an information exchange occurs* at a meeting the belief is updated according to the rule

$$p_{t+1}^i = \mu \tilde{p}_t^i + (1-\mu)\left[X_t^j P + \left(1 - X_t^j\right)\tilde{p}_t^j\right]$$

- This can also be written

$$p_{t+1}^i = \begin{cases} \mu \tilde{p}_t^i + (1-\mu) P, & \text{if } j \text{ audited at } t \\ \mu \tilde{p}_t^i + (1-\mu)\tilde{p}_t^j, & \text{otherwise} \end{cases}$$

- The belief p_{t+1}^i is carried into the next period

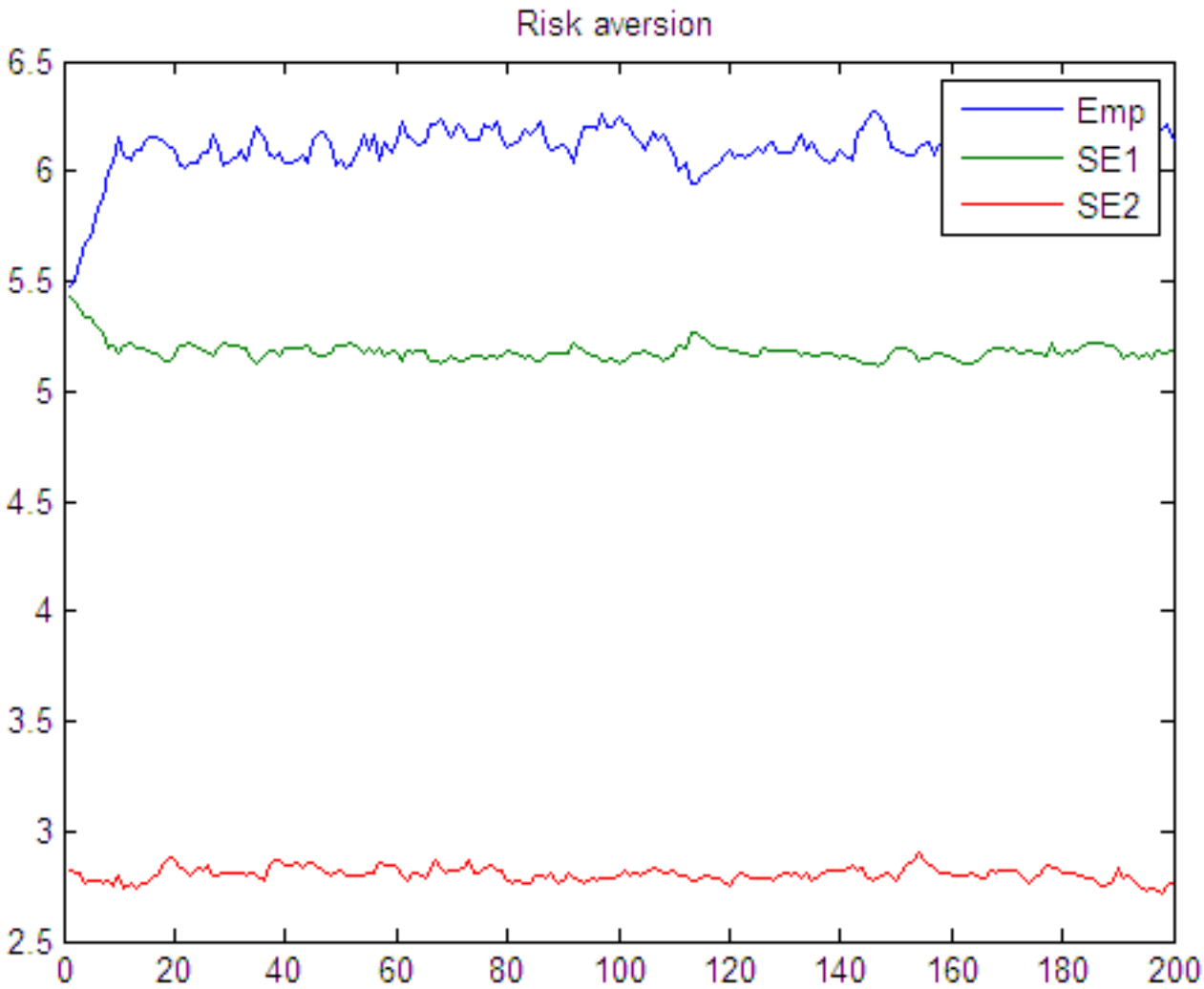
Attitudes

- 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

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

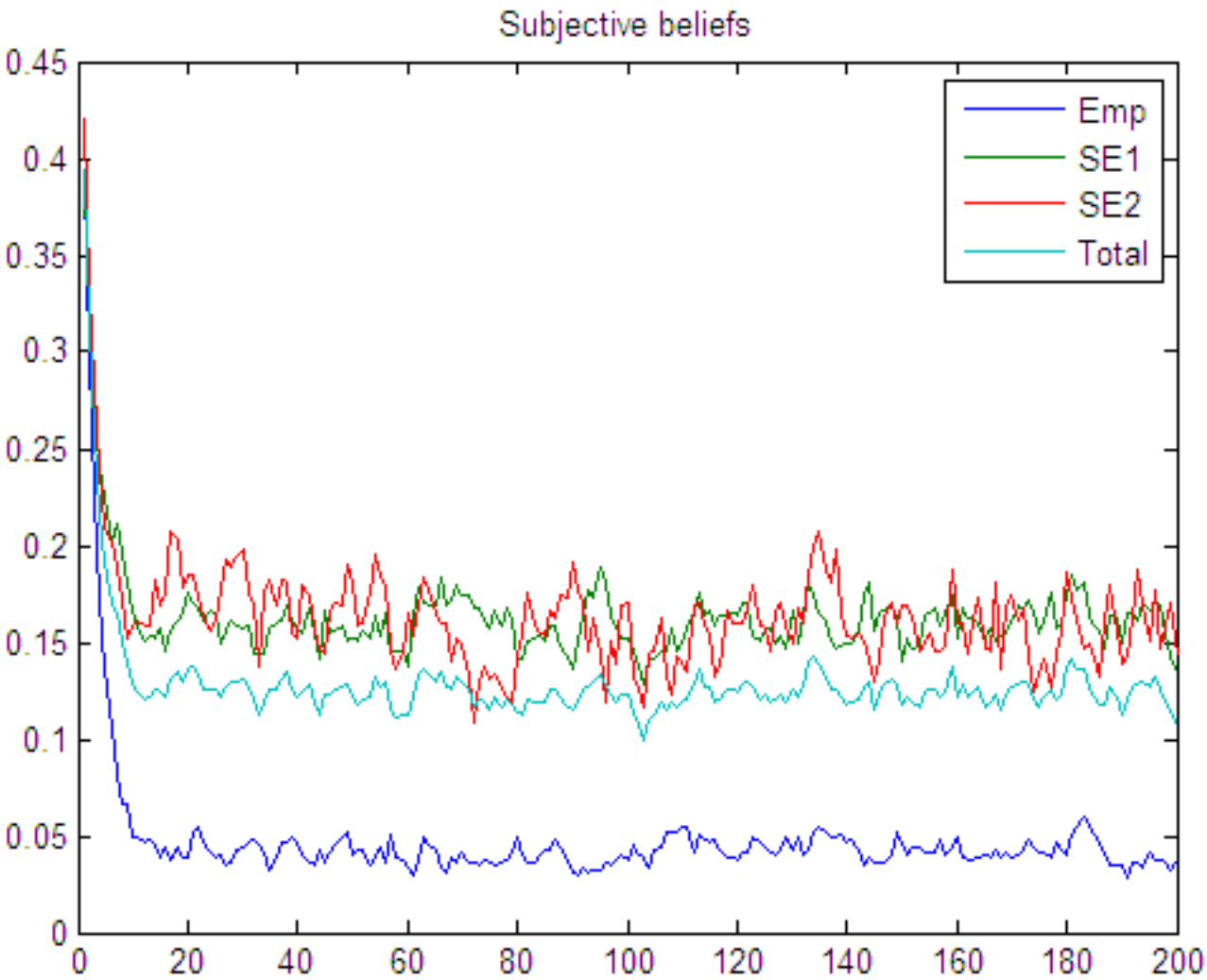
  where $X(i)$ is the number of previous meetings for i at which information was exchanged
- $\chi_{t+1}^i > \chi_t^i$ if information is exchanged with an honest taxpayer and $\chi_{t+1}^i < \chi_t^i$ if information is exchanged with an evader

Risk Aversion

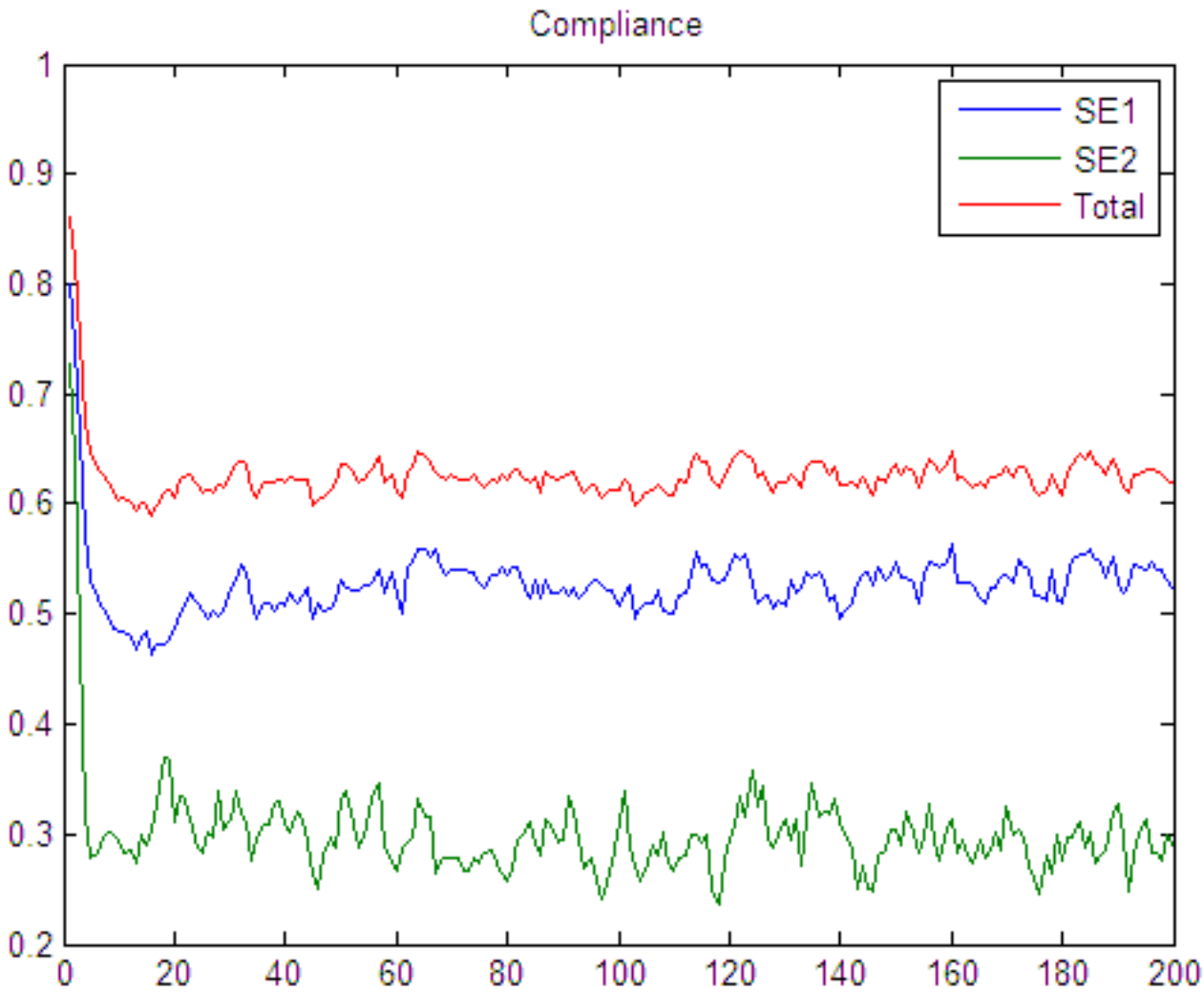


Subjective Beliefs

Audit probability = 0.05



Compliance


Compliance
SE1
SE2
Total
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0
20
40
60
80
100
120
140
160
180
200

Audit Strategies

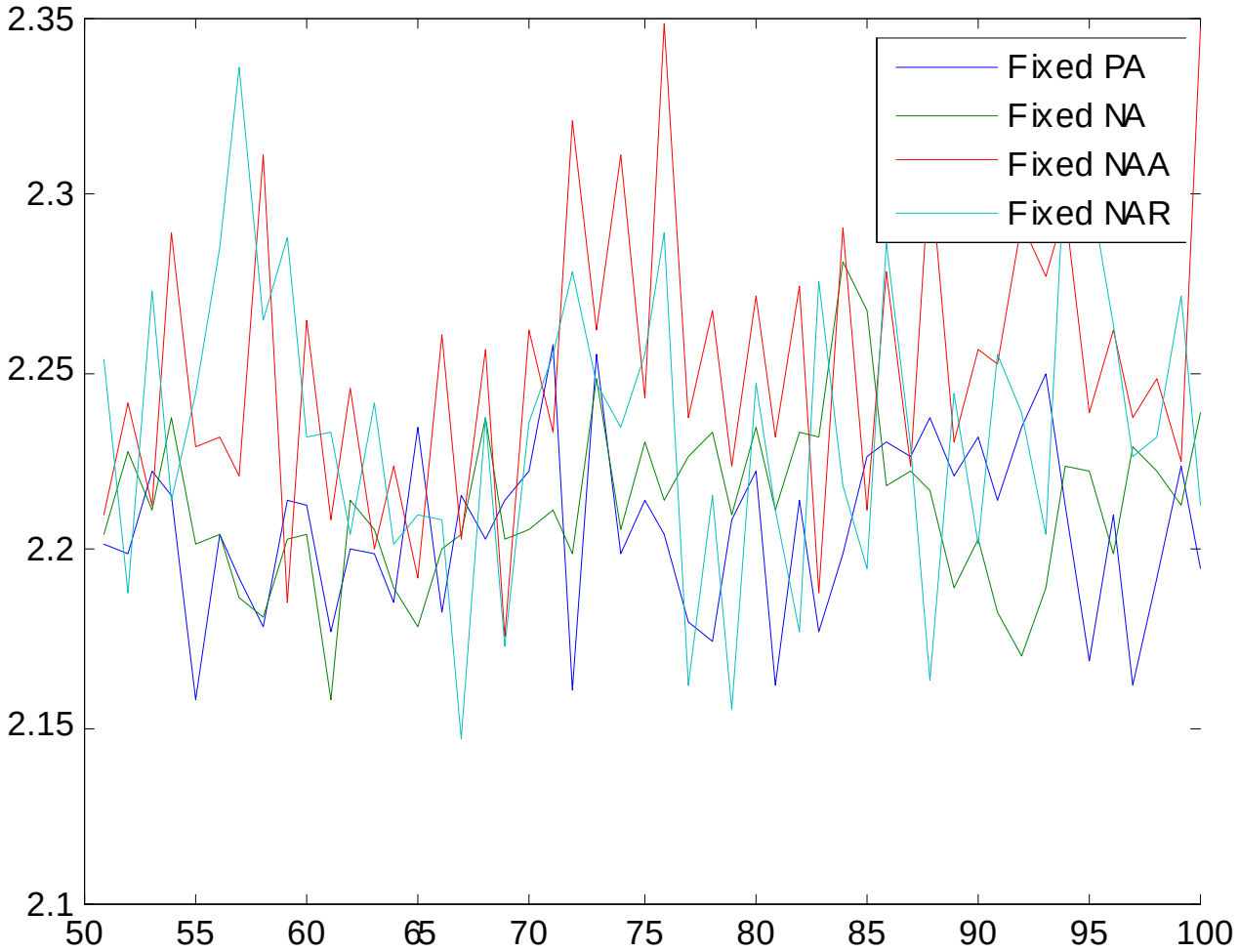
- The model is sufficiently rich to permit a range of questions to be investigated
- We have considered
  - The optimal number of random audits
  - Alternative audit strategies
  - The choice between audit types (hard or soft)
- The focus here will be on alternative audit strategies

Audit Strategies

- Four audit strategies are analyzed:
  - FixedPA: Random audit of the self-employed with a fixed probability
  - FixedNA: Audit a fixed number of taxpayers in each occupation
  - FixedNAA: Switches audits between occupations each period
  - FixedNAR: Randomly switches audits between occupations
- The fixed numbers match the expected number from the random audit

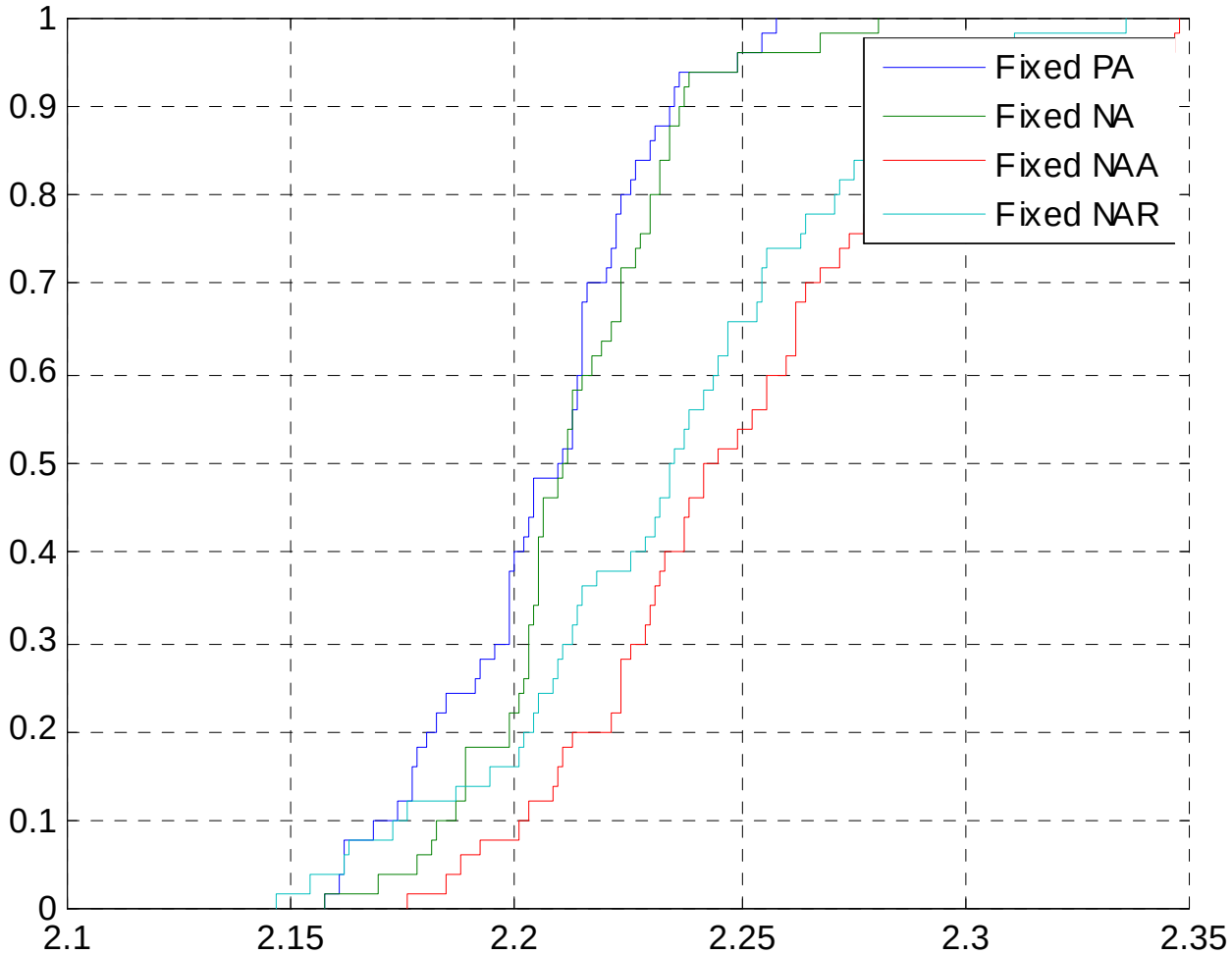
Audit Strategies

Tax and Fine Revenues



Audit Strategies

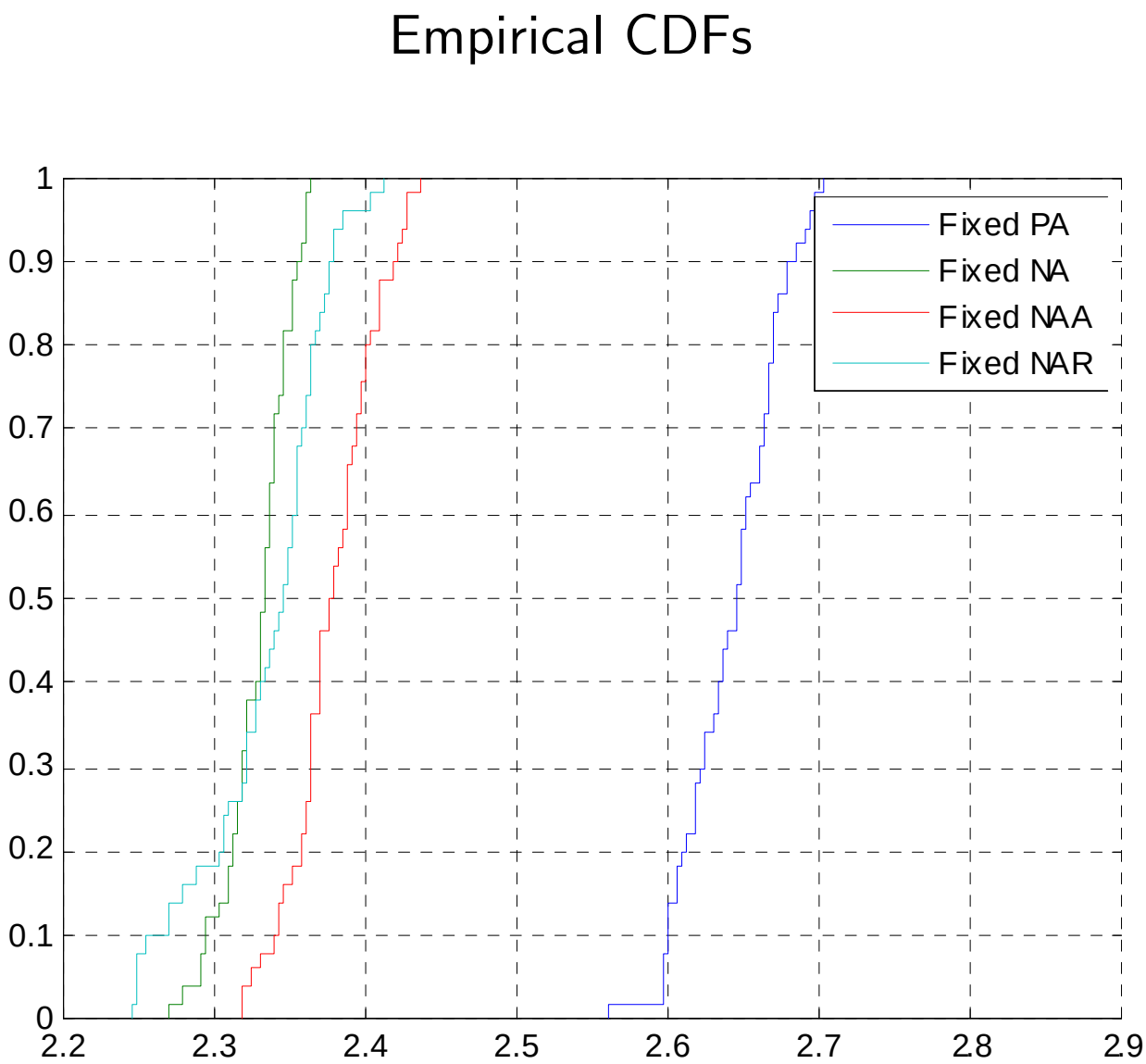
Empirical CDFs



Dual Probabilities

- An extension to the model is to allow each taxpayer to have separate beliefs about the audit probability in the two occupations
- Beliefs are now the pair $\left\{p_t^i(1), p_t^i(2)\right\}$
- If the taxpayer works in occupation j then $p_t^i(j)$ adjusts as before
- In contrast $p_t^i(j')$ only adjusts if information is exchanged with someone working in occupation j'

Dual Probabilities



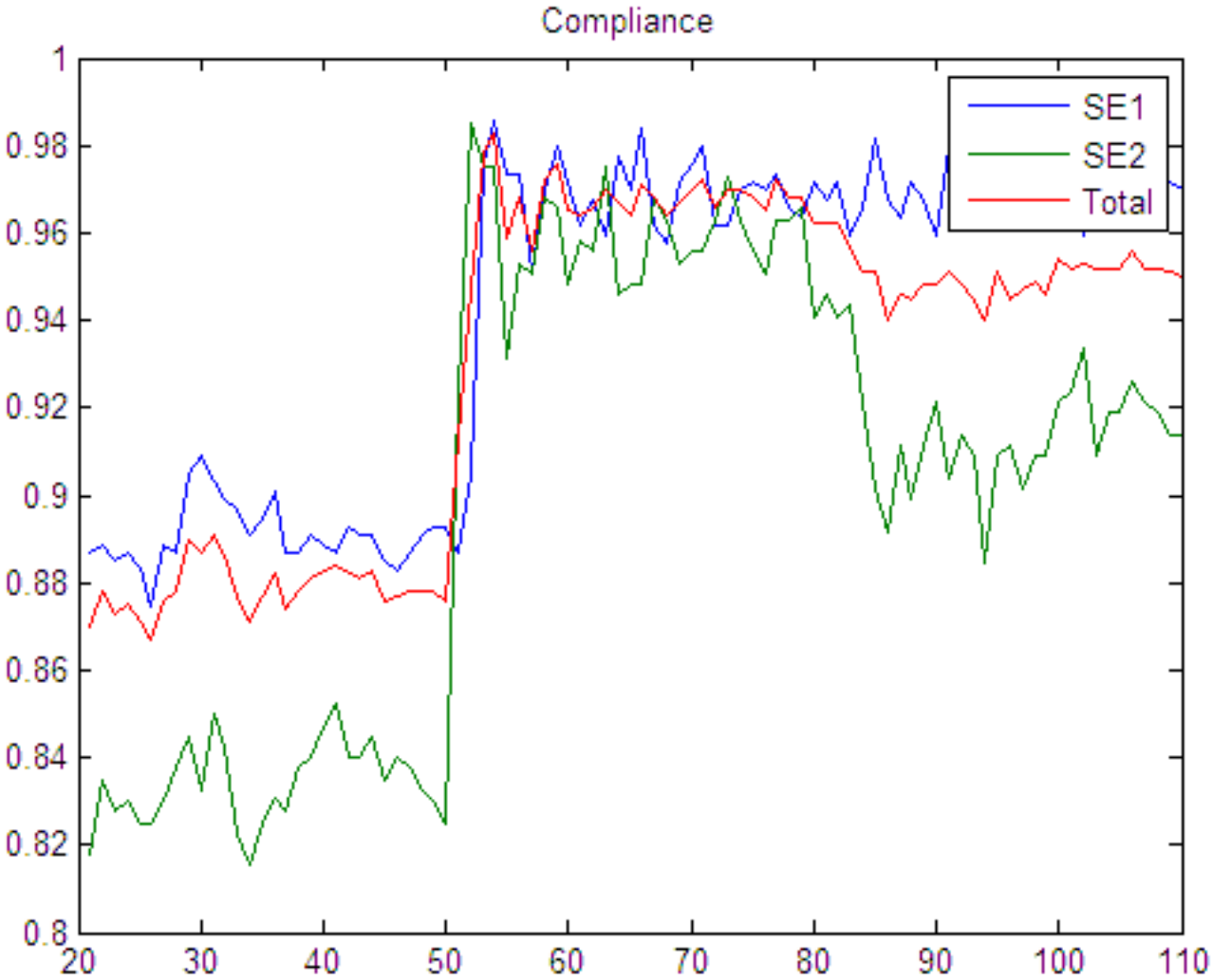
Predictive Analytics

- The role of *predictive analytics* is to identify the best audit targets
- 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

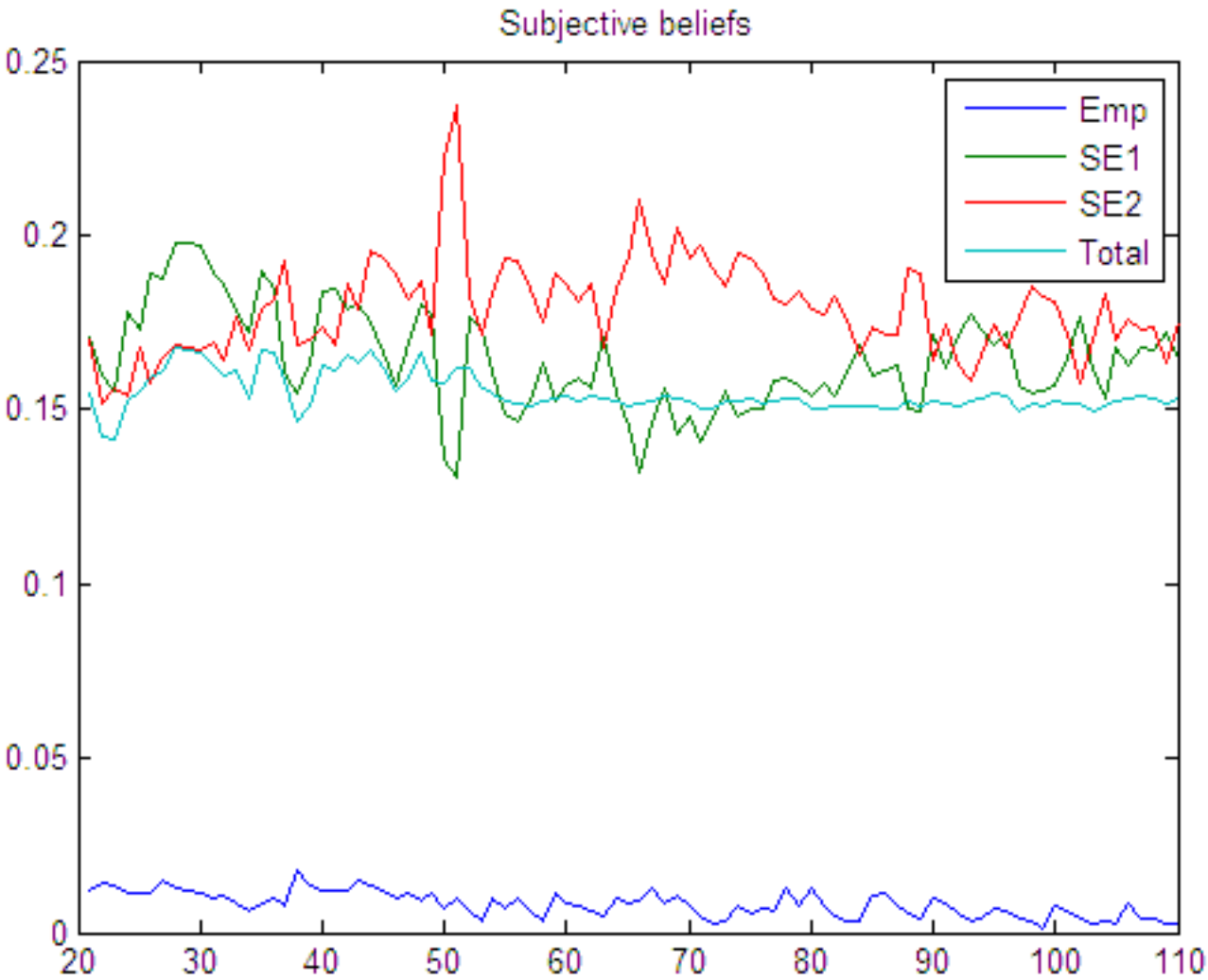
Predictive Analytics

- The simulation uses random audits for the first 50 periods
- The data from audits 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 51-80 the top 5 percent are audited and audit outcomes used to update regression
- For periods 81-110 the top 2.5 percent are audited and 2.5 percent are randomly audited

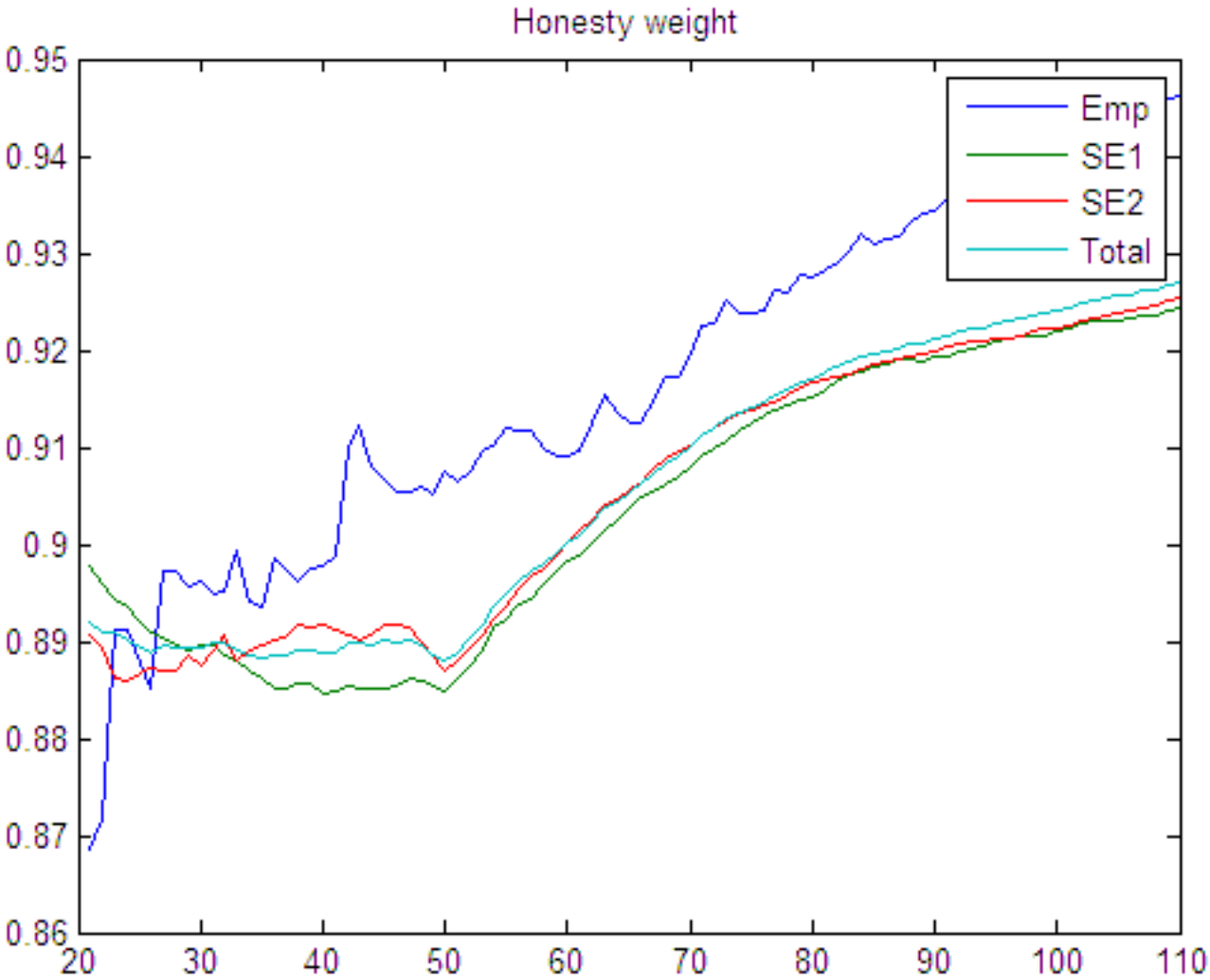
Predictive Analytics



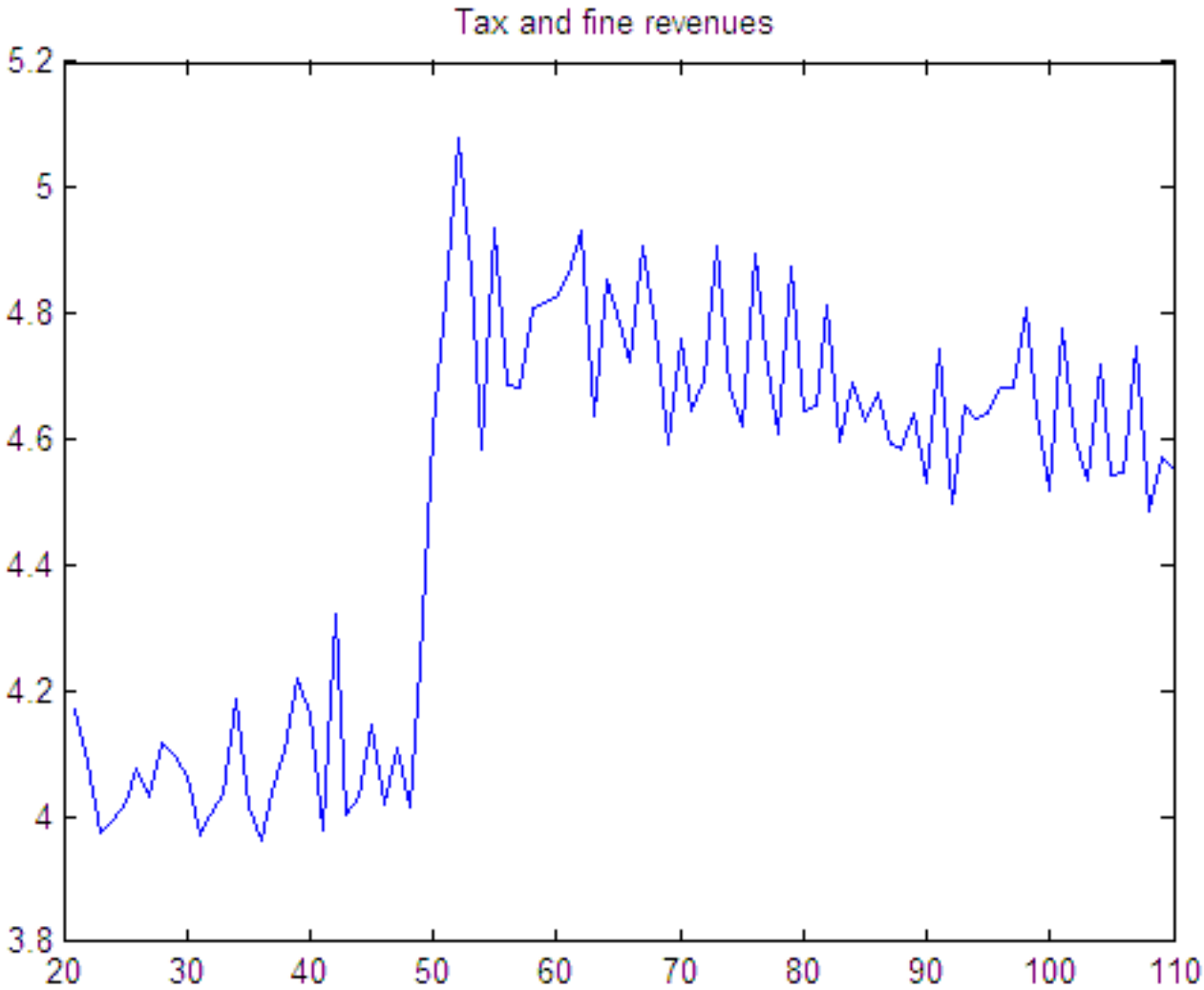
Predictive Analytics



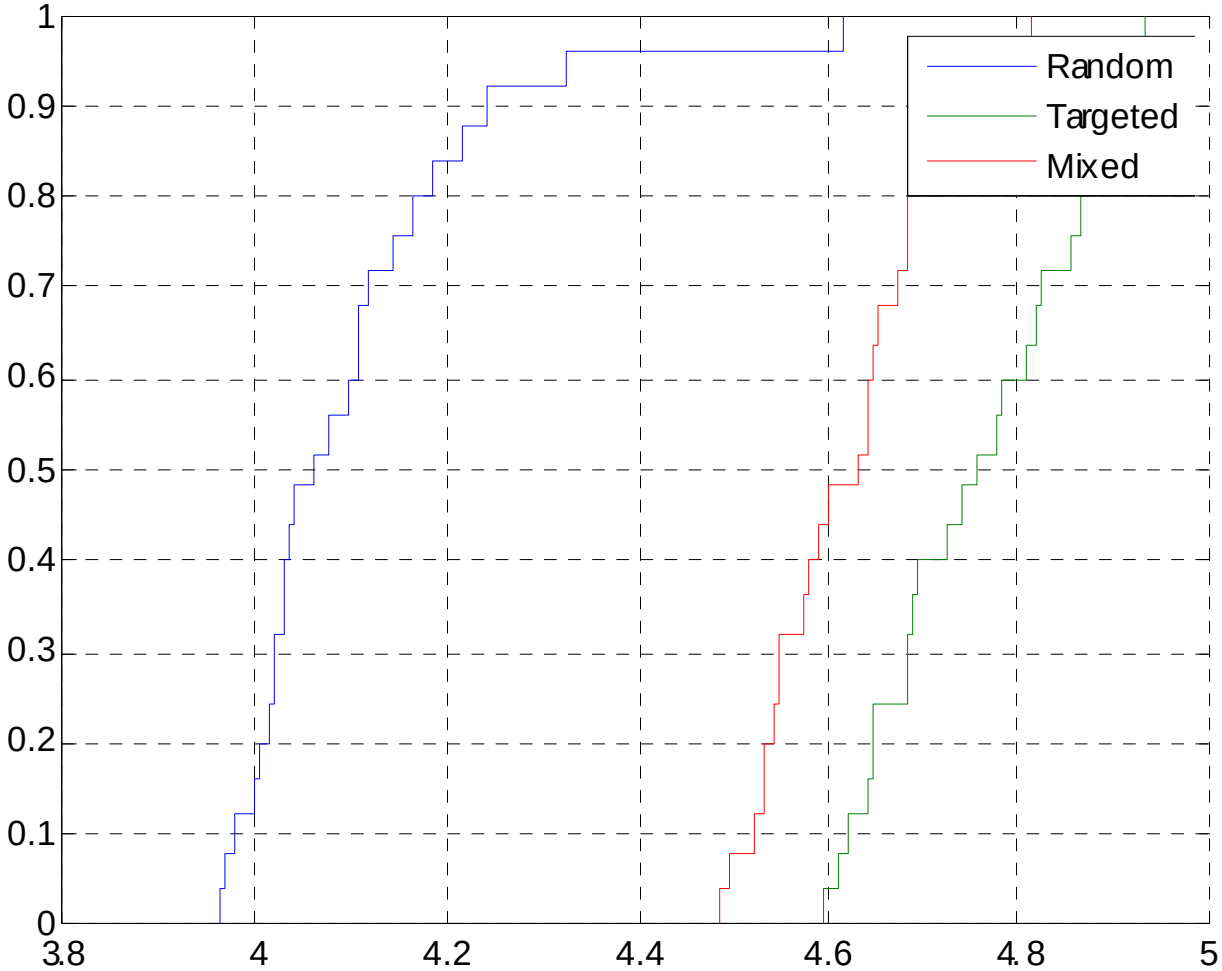
Predictive Analytics



Predictive Analytics



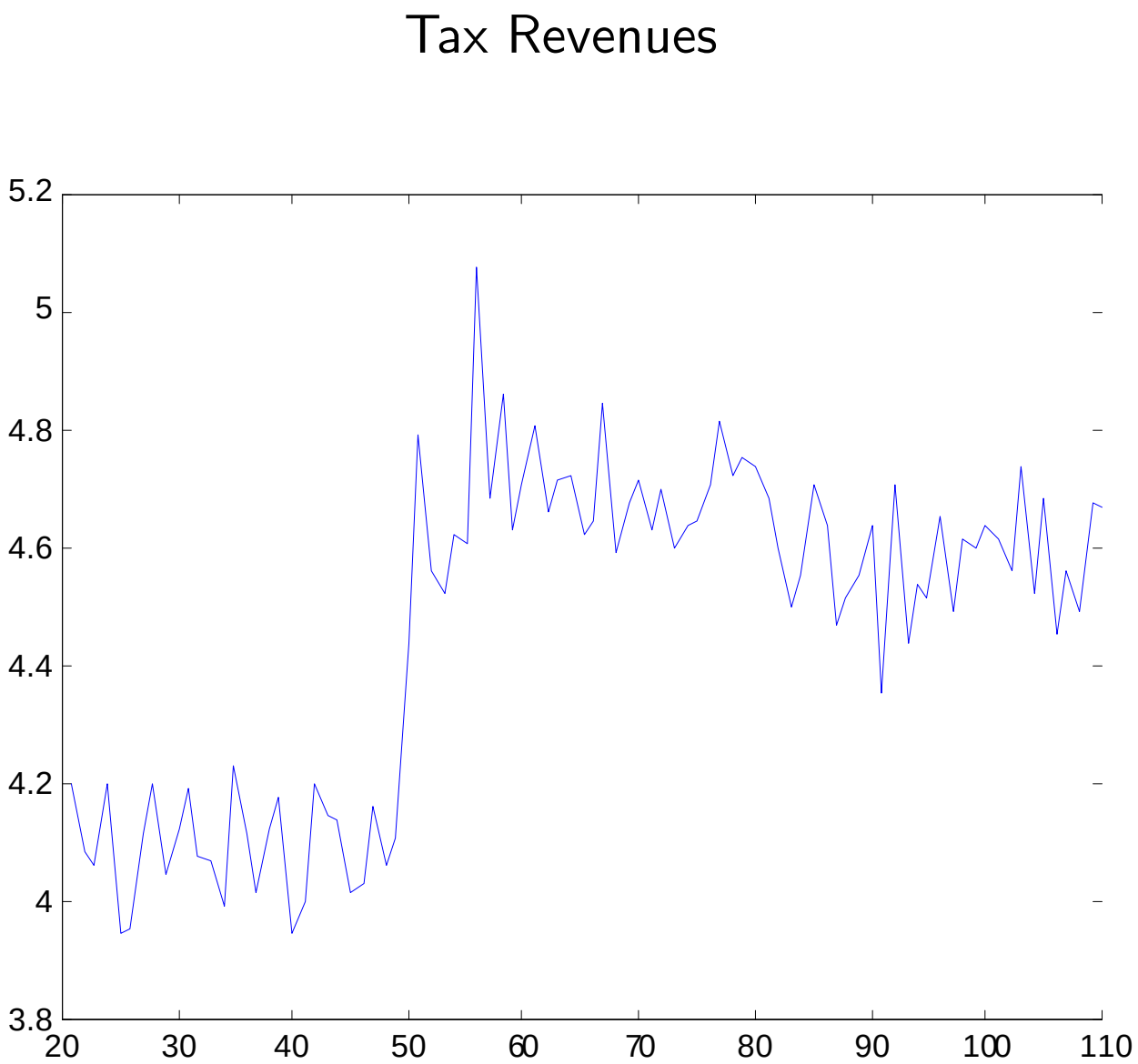
Predictive Analytics



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

Predictive Analytics with Dual Probabilities



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 behaviour can vary significantly across occupational groups
- Predictive analytics is successful in encouraging compliance and increasing revenue