

# Calibration Analysis

September 5, 2017

This document lays out all the plots that we want to use in the report.

## 1 Calibration Analysis

After running all the Latin Hypercube Sampled configurations, we calculated an objective score for each of the configurations to see which ones were better reflecting the real parameters.

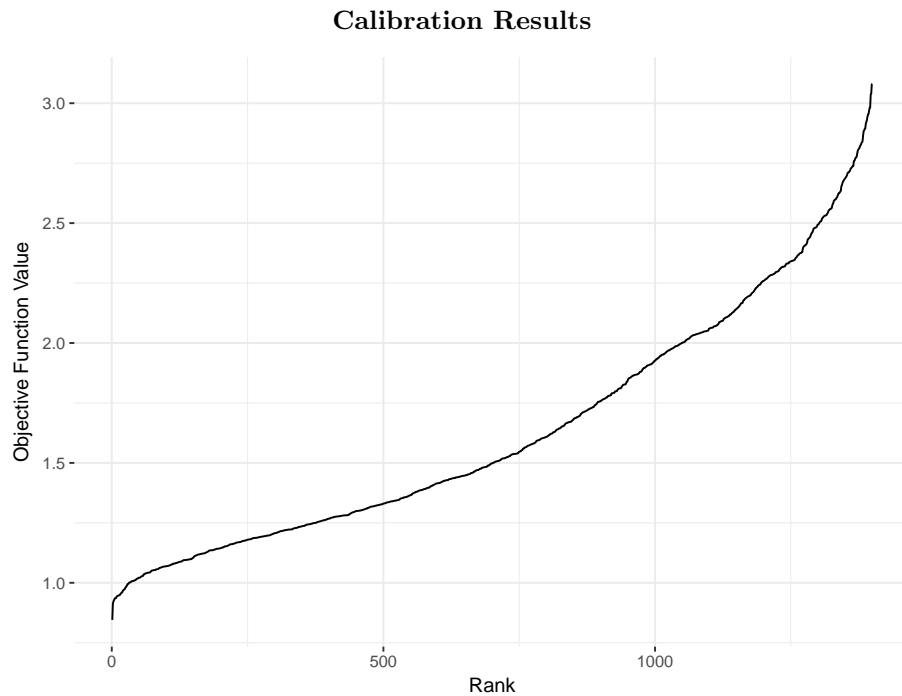


Figure 1: Model-Runs Ranked based on their Objective Function Values

## 2 Sensitivity Results

For the sensitivity analysis, we used Classification and Regression Trees (CART), Recursive Partitioning and Regression Trees (RPART), and RandomForest techniques. In the analysis of each output against the inputs, we ran analysis of variance (ANOVA) using each technique, and then averaged the resulting values. This average is the composite score on which the following heatmaps are produced.

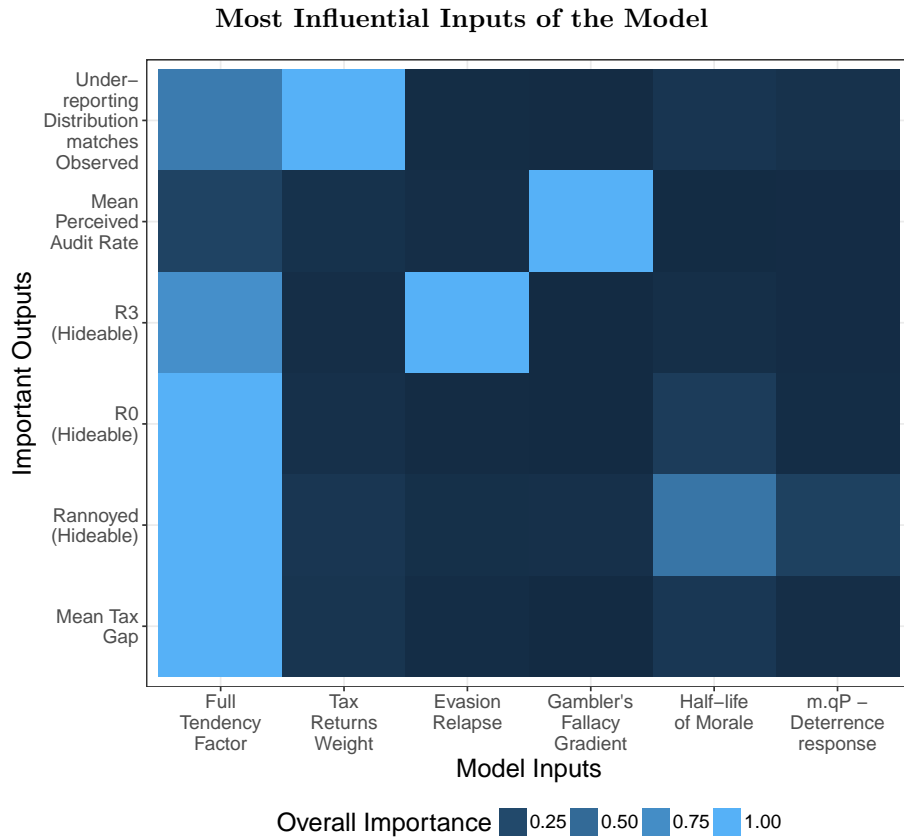


Figure 2: Heatmap of all Important Outputs against their Most Influential Inputs

### 2.1 Some Tree Outputs

These RPART plots show what percentage of model-runs are divide the variance of model-outputs in two halves based on the model-inputs at each level. The figure in the middle shows the average predicted value of the output of interest.

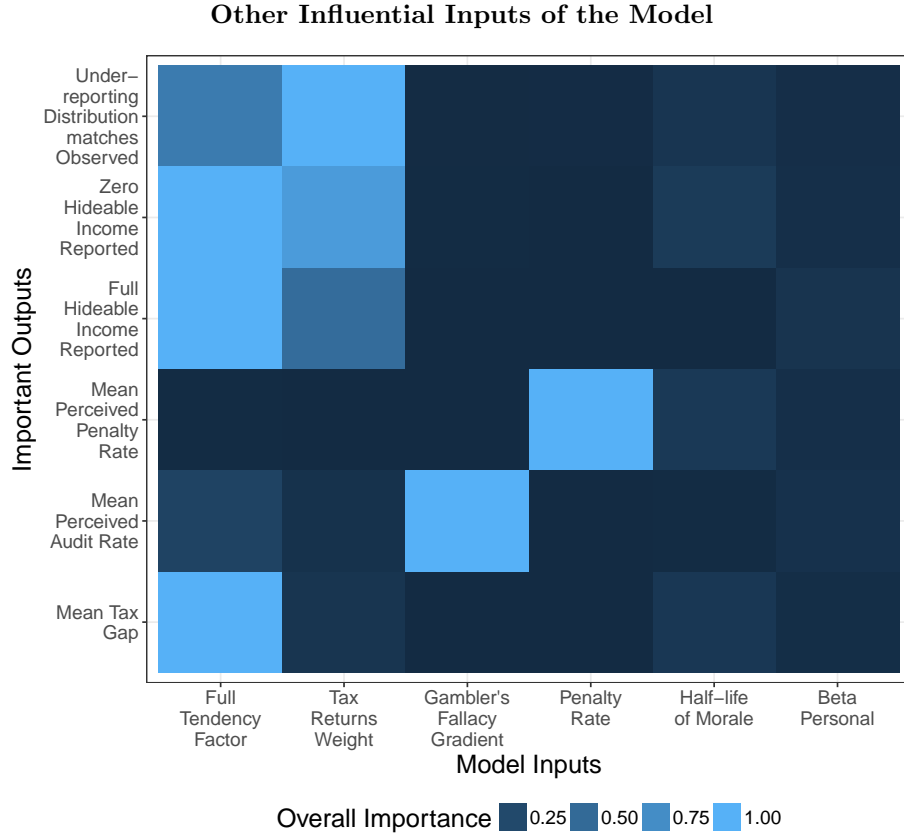


Figure 3: Heatmap of other Important Outputs against their Most Influential Inputs

### 3 Best Case Analysis

Based on the calibration exercise, we ranked all the model-runs to see which model-run closest reproduces the known parameters. After that was done, we took the best case and here are some interesting outputs of the best case.

#### 3.1 Aggregate Level Outputs

Aggregate level outputs are perhaps among the most important outputs of interest for decision makers. The following figures show how tax gap, perceived audit rate, and perceived penalty rate of the simulated population, as a whole, changes over time. Unless otherwise mentioned, all income tax evasion are expressed in terms of hideable income rather than actual income of agents. Hideable income is the percentage of an individual's income that is not expressly known to the tax

authorities. Self employed individuals, in general, have a higher hideable income than non-self-employed individuals due to appropriate reporting mechanisms in place.

### **3.2 Individual Level Outputs**

While the above plots were aggregate level outputs, the more interesting plots are the ones that show individual trajectories. Here are a few plots showing the behavior of individuals grouped by their employment status (self-employed or non-self-employed). Self-employed individuals are known to have more opportunities to evade taxes, than non-self-employed people. These plots show how much of the hideable income to agents report.

### **3.3 Population Distributions**

Now that we have seen aggregate level outputs and individual behaviors, here are a few plots examining the distribution of outputs among the simulated population.

## Variance of Mean Percieved Audit Rate

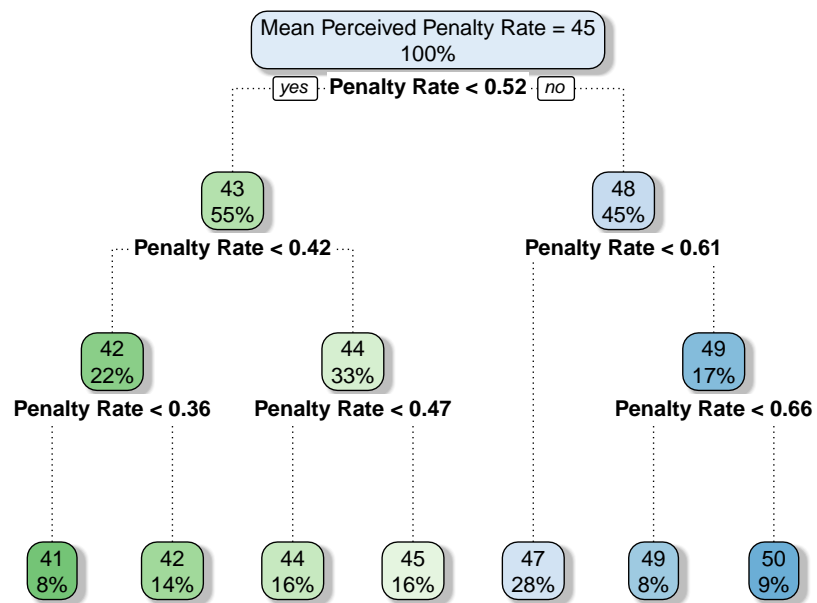


Figure 4: RPART Plot of Mean Perceived Penalty Rate

## Variance of Tax Gap

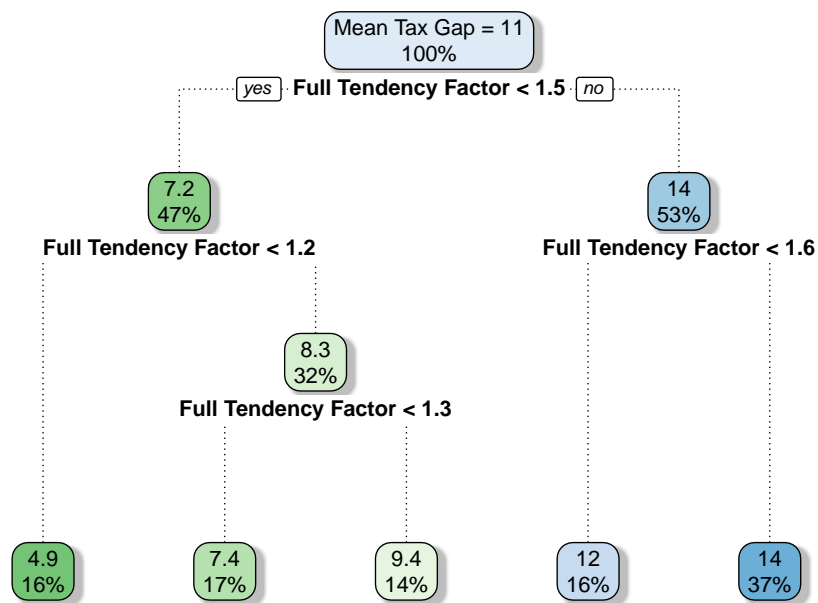


Figure 5: RPART Plot of Mean Tax Gap

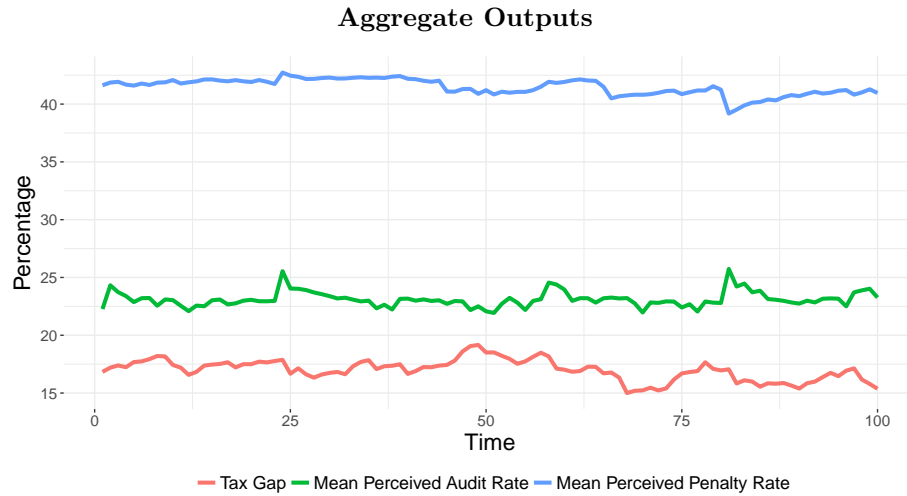


Figure 6: Aggregate Outputs of the Best Case

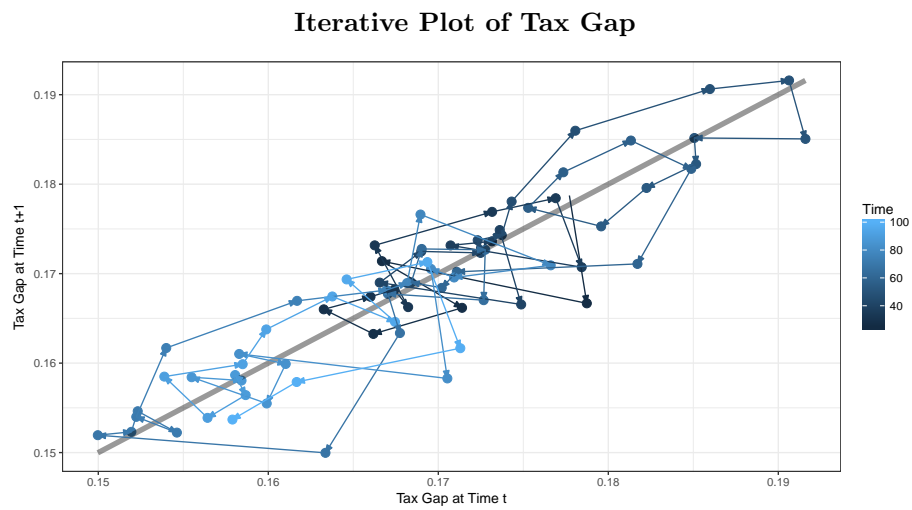


Figure 7: Iterative Plot Showing how Tax Gap Varies from One Year to the Next

Selected Trajectories of Self-Employed Agents

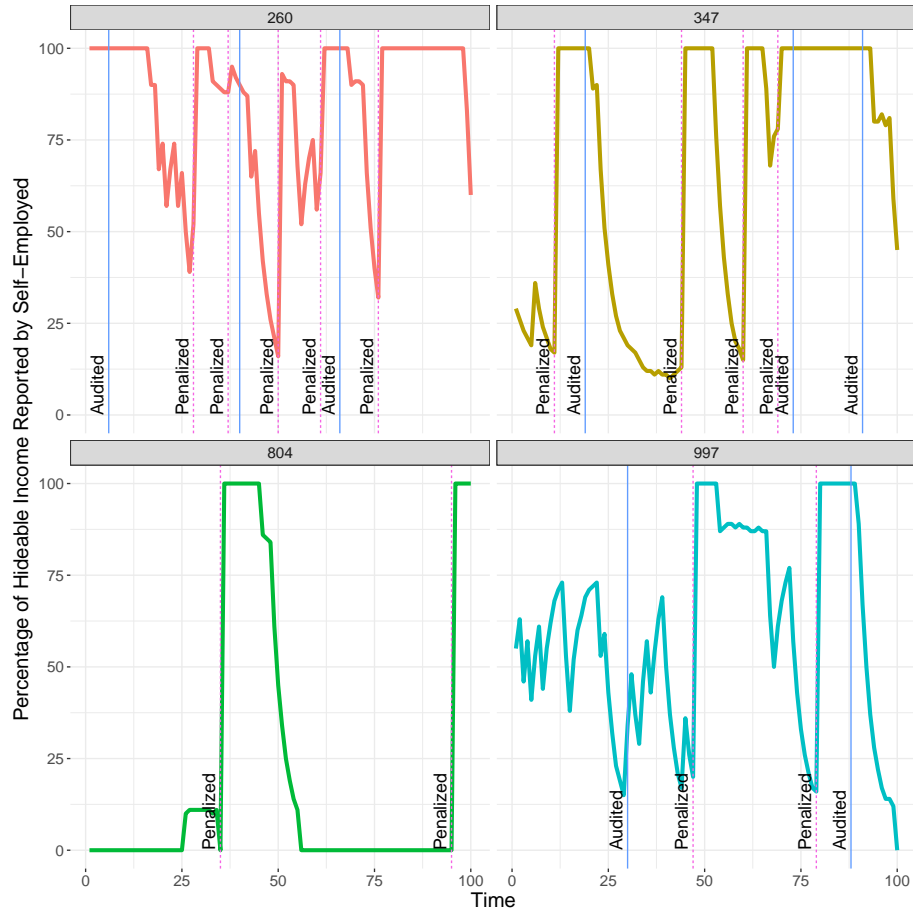


Figure 8: Individual Trajectories showing Hideable Income Reported by a few Self-Employed Agents



### Selected Trajectories of Non-Self-Employed Agents

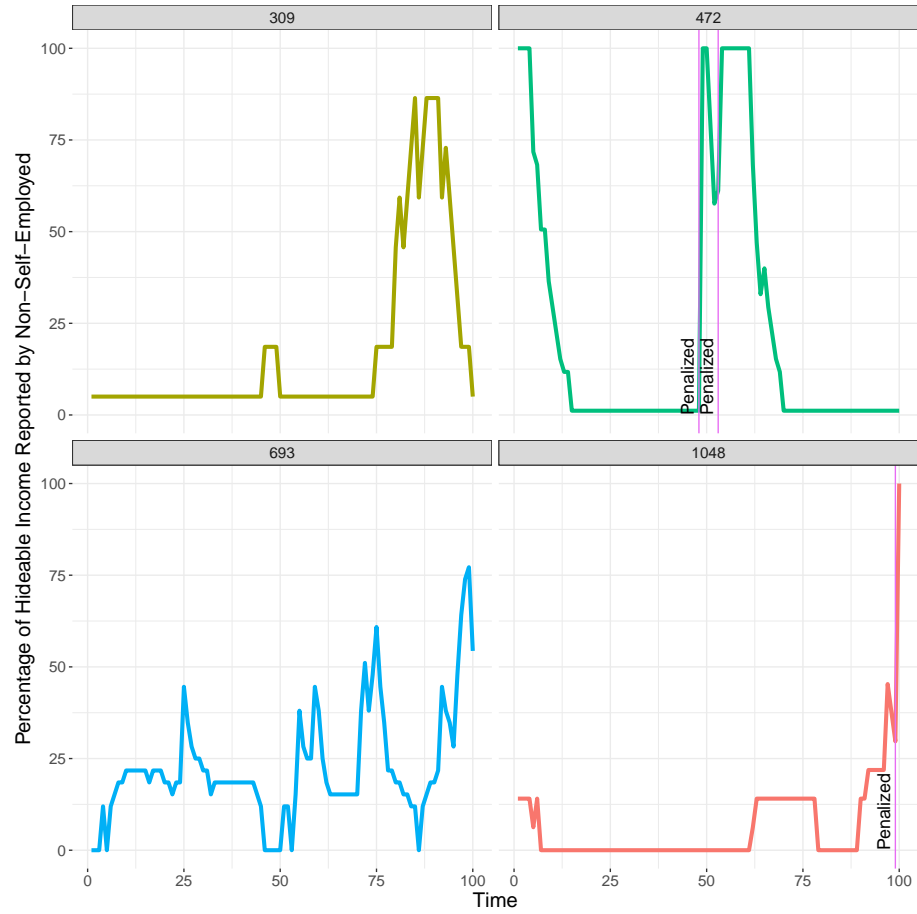


Figure 9: Individual Trajectories showing Hideable Income Reported by a few Non-Self-Employed Agents

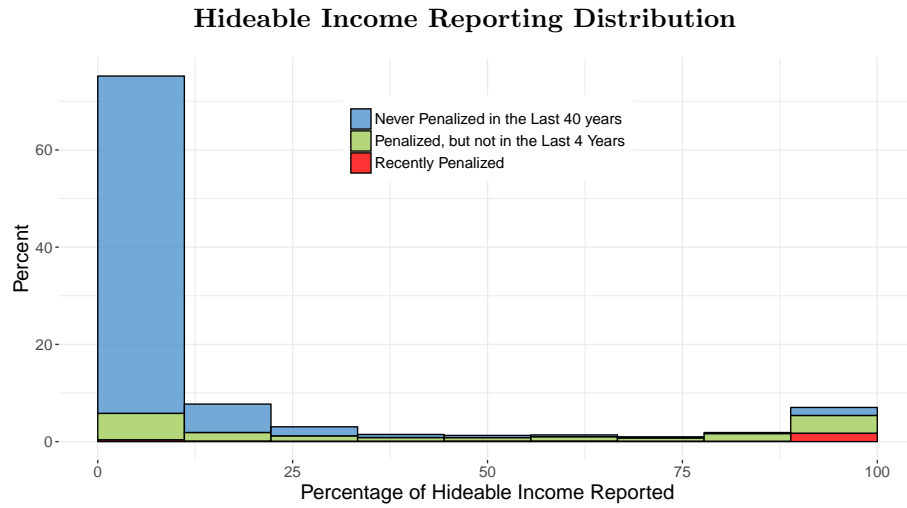


Figure 10: Distribution of how Hideable Income is Reported by Agents

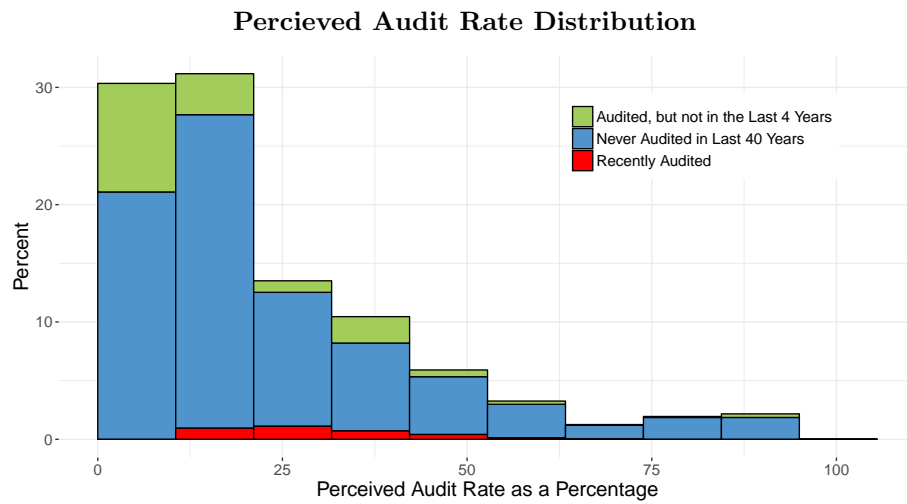


Figure 11: Distribution of Perceived Audit Rate of Agents



Figure 12: Distribution of Perceived Penalty Rate of Agents

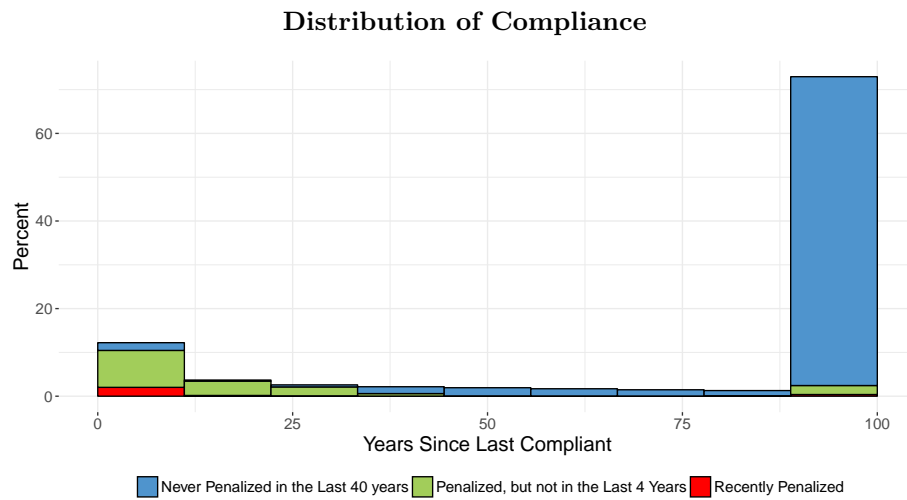


Figure 13: Distribution of Compliance Behavior of Agents Over Time

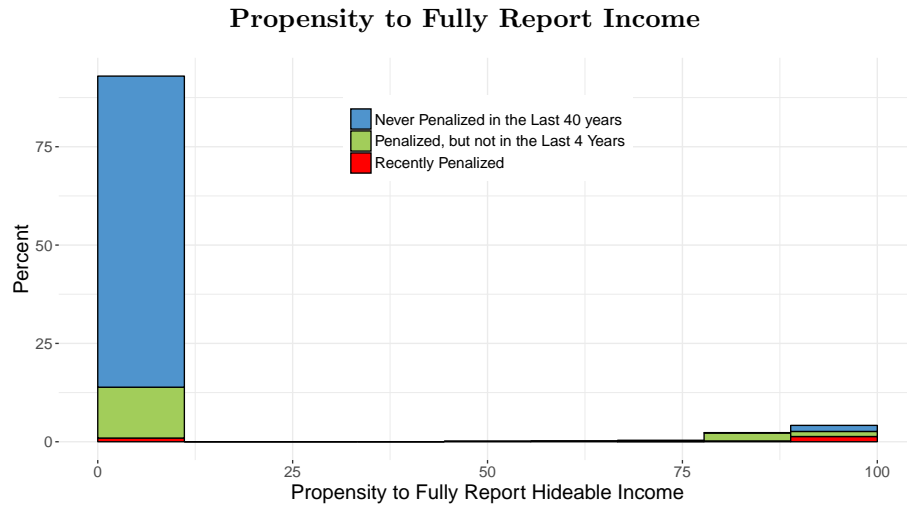


Figure 14: Distribution showing the Propensity of Agents to Report All of their Income

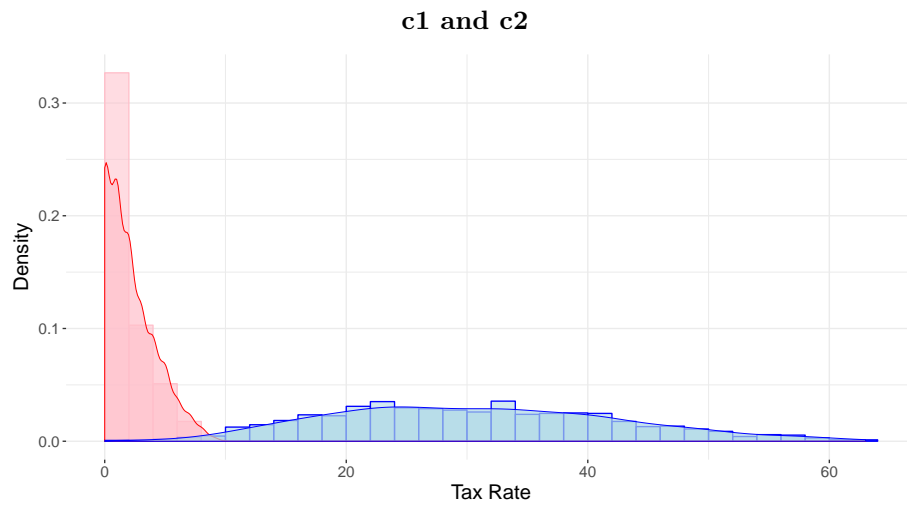


Figure 15: Distribution of c1 and c2 Thresholds of the Agents