RAND's Interdisciplinary Behavioral and Social Science Agent-Based Model of Income Tax Evasion

Technical Report

Raffaele Vardavas, Pavan Katkar, Andrew M. Parker, Gursel Rafig oglu Aliyev, Marlon Graf, Krishna Kumar

RAND Education and Labor

WR-1233
September 2019

RAND working papers are intended to share researchers’ latest findings and to solicit informal peer review. They have been approved for circulation by RAND Education and Labor but have not been formally edited or peer reviewed. Unless otherwise indicated, working papers can be quoted and cited without permission of the author, provided the source is clearly referred to as a working paper. RAND’s publications do not necessarily reflect the opinions of its research clients and sponsors. RAND® is a registered trademark.
Abstract: Income tax evasion is a problem that poses considerable challenges for tax authorities and governments at the local, state, federal levels, as well as internationally. Its causes and implications are both economic and social, and therefore it is of enormous importance in policy design. We built an agent-based computational simulation model of income tax evasion. Within the simulation, individuals’ compliance behavior changes through an adaptation process based on their past experiences with audits and tax evasion penalties, their perception of the fairness in taxation rates and social interactions with people in their social networks. To inform the model we have conducted a survey on a nationally representative sample on the perceptions of tax fairness. The specific purpose of our survey was to guide the model construction, test our model assumptions, as well as inform behavioral parameter values and the calibration procedure. In addition, the survey provides novel insights into the social dynamics of risk and fairness perceptions, including how they are influenced by perceptions and experiences of social network contacts and in community at large.

Here, we present technical details that describe our model and how it was informed by our survey. In our first two sections we provide a brief introduction to the problem and an overview of past agent-based models of tax compliance. Sections 3 to 11 provide a description of our agent based model following the overview, design concepts, and details protocol [69, 70]. Section 12 to 15 provides description of our survey and focuses only on the analyses that helped inform the model and in particular the behavioral mechanisms and parameter values. Sections 17 to 20 described model verification, validation and calibration. Sections 21 and 22 describes results from possible intervention policy. Finally, Sections 23–24 provide a discussion of our results and describe limitations and future work.

Acknowledgements: We gratefully thank the National Science Foundation’s Interdisciplinary Behavioral and Social Science (IBSS) Research program for their support of this research. Award Number 1519116. We are particularly grateful to Dr. Kim M. Bloomquist for his assistance in informing the model and the survey design and to Drs. James Anderson and Robert Lempert for consultations. We further thank the RAND ALP team including Ms. Karen Edwards, Mr. Alerk Amin, Mr. Sherban Drulea, Dr. David Grant, Dr. David Kennedy and Dr. Tim Colvin.

Plagiarism: To the best of our knowledge we have acknowledged the work and the descriptions provided by others. We have used the services provided by quetext.com to check for any unintentional plagiarism.
1 Introduction

2 Understanding social drivers of tax evasion
  2.1 Shaping perceptions on tax evasion
  2.2 Media’s impact on individual-level tax compliance

3 Overview of ABMs and Past Models of Tax Compliance
  3.1 Complex Adaptive Systems and Agent-Based Models
  3.2 A Brief Overview of Past Models of Tax Compliance

4 Purpose

5 Entities, State variables, and Scales
  5.1 The Taxpayers
  5.2 The Internal Revenue Service (IRS)
  5.3 The Governing State (GS)
  5.4 The Media

6 Model Overview and Scheduling
  6.1 General Overview
  6.2 The Overview of the Taxpayer’s Behavioral Model (TBM)
  6.3 The Model Code

7 Specifications of the Taxpayer’s Behavioral Model Components
  7.1 The calculation for the taxpayer’s personal evaluation $\Delta_p^{(i)}$
  7.2 The calculation of the effect of the penalties $\Delta_{\text{penalty}}^{(i)}$
  7.3 The calculation of the media feedback term on tax morale and fairness $\Delta_M^{(i)}$
  7.4 The effect of the media feedback on the taxpayer’s perceived audit rate $\hat{a}_t^{(i)}$
  7.5 The calculation for the taxpayer’s perceived audit rate $\hat{q}_t^{(i)}$
  7.6 The calculation for the taxpayer’s perceived penalty rate $\hat{P}_t^{(i)}$

8 Specifications and Initialization of the IRS Model Components
  8.1 The effective tax rate
  8.2 The transitions between a tax refund and balance due
  8.3 The initial IRS audit strategy and rate
  8.4 IRS Fixed Operational Costs and Variable Audit Costs

9 Specifications of the Governing State (GS) Components
  9.1 The total revenues to the GS
  9.2 The Tax Gap
  9.3 Scaling of the income tax revenues
  9.4 The total costs to the GS
  9.5 The marginal budget deficit
  9.6 The fiscal policy of the GS
  9.7 Estimating the changes in the US Government Debt to GDP ratio
19 Uncertainty and Sensitivity Analyses .......................... 122
  19.1 Uncertainty Analysis ........................................ 122
  19.2 Sensitivity Analysis ........................................... 123

20 Model Calibration .................................................. 129
  20.1 Results in Matching our Calibration Targets .............. 130
  20.2 Analysis of the correlations in our model input parameters due to Calibration ................................. 134

21 Simple intervention and "what if" scenarios ....................... 136
  21.1 "What if" scenario with more equitable income distribution ........................................ 136
  21.2 Changes in the proportion of self-employed taxpayers ............................................... 138
  21.3 Changes resulting from the 2018 Tax Reform .................... 141
  21.4 Cost effectiveness in direct benefits of increasing the audit rate .................................. 143
  21.5 Cost effectiveness in indirect benefits of increasing the audit rate .................................. 146
  21.6 Intermittent changes in the audit rate ........................................ 148

22 Combining fiscal and deterrence intervention policies .......... 149
  22.1 Part 1: Increasing the voluntary compliance rate ......... 150
  22.2 Part 2: Maintaining the higher compliance rate and reducing the marginal debt ...................... 154

23 Summary and Discussion ........................................... 158
  23.1 Summary of our survey design and analyses and secondary data used ................................. 158
  23.2 Summary of model verification, validation, and calibration .............................................. 160
  23.3 Summary of hypothetical "what-if" scenario experiments .................................................. 161
  23.4 Summary of tax reform policies and changes in the audit rate ........................................... 162
  23.5 Summary of policies that combine changes in tax and audit rates ..................................... 163

24 Limitations ............................................................ 163
  24.1 Experience Discounting and the EWMA process ......... 166
  24.2 Eliciting social influence parameters for tax compliance behavior using the results from Alm et al. 2009 ........................................ 171
  24.3 Program Evaluation and Review Technique (PERT) distribution ........................................... 176
  24.4 Survey Questions of the American Life Panel Well Being 456 .............................................. 179
  24.5 Fitting data to an S-Curve ........................................... 187
  24.6 Estimating Tax Rate, Audit Rate, Penalty Rate Elasticities and the S-Curve Parameters ......... 188
Author’s Summary

Income tax evasion is a problem that poses considerable challenges for tax authorities and governments at the local, state, federal levels, as well as internationally. Its causes and implications are both economic and social, and therefore it is of enormous importance in policy design. Agent-based models, a type of computer simulation models, can be used as tools to understand how taxpayers change their compliance behavior and what government policies could help mitigate the problem of tax evasion. These models can be run thousands of times to map possible futures under different assumptions about taxpayer’s behaviors and how they interact with each other and with new environments they help form. Within the model, we can adjust policy levers and explore different “what-if” scenarios to see how tax compliance outcomes are affected at the aggregate-level.

In this project, we built an agent-based computational simulation model of individual income tax evasion. Within the simulation, taxpayers’ compliance behavior changes through an adaptation process based on their past experiences with audits and penalties, their perception of the fairness in taxation rates, interactions with people in their social networks and on what the media reports about the levels of tax evasion. The model was designed to explore policies that could lead to amplification effects and potentially to societal transformations regarding tax compliance. We were particularly interested in policy experiments that combined increases in IRS audit rates with decreases in tax rates to explore the dynamics of both the overall compliance rate and of the additional debt due to the potential reductions in income tax revenues. Our aim was to find policies that can achieve significant increases in compliance while leading to economically manageable debts over the years.

To inform the model we have conducted a tailored behavioral and social network survey on a nationally representative sample on the perceptions of tax fairness. The specific purpose of our survey was to guide the model development and inform behavioral parameter values including taxpayers’ perceptions of tax fairness and the risk of being audited. The social network survey informed the model on how self-employed and salaried taxpayers with different experiences with audits mix socially, and provided insights into how tax-related perceptions are influenced by social contacts and in the community at large. In addition, our model was informed from multiple data sources including census data, IRS data and from a dataset providing a US representative synthetic population of how residents connect in large urban social networks.

Model validation and calibration were based on reproducing known stylized facts and summary statistics describing taxpayers’ compliance behavior and their aggregated outcome that characterize US settings for the year 2016. In particular, the model reproduces the U-shaped distribution describing individual-level compliance behavior of the proportion of income that can be hidden to the IRS, and the tax gaps (i.e., the ratio between the net underreported tax and the total true tax liability) with an interquartile range of 12.0% to 15.5%. Using the IRS tables that specify audit rates by income groups, our model further reproduces the average costs of an IRS audit.

We carried out a set of experiments that considered different "what-if" scenarios. For these experiments, our model suggests that (i) a more equitable income distributions of taxpayers could lead to an increase in the tax gap by between 0% to 2%, (ii) increasing the audit rate from just below 1% to 2.8% could lead to a tax gap ranging between 5.8% and 7.9%, and the increase in audit costs would be balanced by additional detected unpaid taxes, and (iii) unless taxpayer’s incomes or audit rates increase, the recent tax reforms would reduce the tax gap by between 0% to 0.5% and could lead to between $228 to $241 billion in less individual income tax revenues. Policies that combine a reduction in tax rates with an increase in the audit rate can lead to significant reductions in the tax gap. However, our model suggests that tax compliance cannot be increased rapidly with a tax reduction unless it is accompanied by an increase in the audit rate. To reduce the tax gap to under 5% within the next 25 years, the audit rate would need to increase to just under 4%. In this scenario, taxpayer’s effective tax rates can be reduced by at most 5% without leading to significant increases in the budget deficit.
Introduction

1 Introduction

Income tax evasion is a global problem that poses considerable challenges for tax authorities and governments. It involves a broad range of issues within economics and the social sciences, and it is of enormous importance in policy design [177]. The problem is especially acute in developing countries, where taxpayer compliance is estimated to be as low as 50%. The problem is also widespread in more developed nations, and in the US, substantial tax evasion persists [157]. The Internal Revenue Service (IRS) estimated that in the fiscal year 2008-2010, Americans owed $458 billion more than they paid or about 18% of federal revenues [82]. This represents an increase of $113 billion compared to their previous estimate for the fiscal year of 2001 [81]. Despite substantial attention to this problem, the factors that drive individuals to evade taxes remain poorly understood.

Researchers have been developing tax evasion models since 1972. These models have attempted to include factors believed to drive tax evasion and thus test different audit strategies that may reduce the number of tax evaders. The vast majority of these models assume rational individuals who use deductive-reasoning to make their tax compliance decisions based on perfect knowledge of audit rates and penalties and occasionally on perfect information regarding the intentions of other individuals. However, in recent years, borrowing from approaches applied to the modeling of other socio-economic systems, a few agent-based tax evasion models have emerged in which these assumptions have been relaxed. In particular, some models have incorporated the insight that individuals often use inductive-reasoning to make their choices by relying on their past experiences [138]. These types of models promise to give better insights into the potential impact of different tax policies may have on discouraging tax evasion [22]. Most models have focused on how individuals respond to policy measures and control options available to the tax authority, in particular, audit rates and penalties for evaders [64].

However, models limited to changes in audit rates and penalty levels as principal policy controls overlook an important aspect of the problem: namely, the relationship between evasion and the perception of tax fairness [52, 89, 103, 121, 174]. Research has shown that people are more likely to evade if they know or perceive that others are evading and not being caught [7, 49, 119, 164, 165, 176]. They fear that they are paying more taxes than they should because others are either evading completely or paying less than they owe. This re-enforcing belief can result in a widespread and social norm of "tax rebellion" that leads otherwise honest taxpayers to feel that the services the state offers are no longer worth the price they are paying [42, 148]. The way that fairness and risk perceptions spread socially, may lead to the formation of a tax evasion or compliance culture. Greece and Italy are two key examples of countries suffering from this attitude resulting in widespread evasion; this, in turn, leads to a shadow economy representing nearly 30% of their GDPs [141, 145, 156]. Clearly, this degree of evasion is a significant problem for governments trying to balance budgets while experiencing these severe social and economic pressures. In contrast, Scandinavian countries such as Denmark can maintain high compliance levels despite high-income taxation levels (above 50%). In these countries, taxes are generally perceived to be fair given the services they provide, and the perception is that others are paying their fair share.

Along with perceptions about tax fairness, attitudes and trust toward government, social and personal norms have been receiving an increased attention as driving forces of the tax compliance in the literature. Kirchler (2007) provides an excellent account of tax studies focusing on these
factors. An example of more recent such studies is a paper by Shu et al. (2012). Through an experimental design, the authors suggest that making ethics salient by signing tax return at the beginning rather than at the end can increase tax compliance. Using the results of another set of experiments, Lamberton et al. (2014) demonstrated that eliciting taxpayer preferences on government spending can also increase tax compliance. In another study that utilized data from a natural field experiment in the UK, Hallsworth et al. (2014) show that tax compliance can be significantly improved by simply reminding taxpayers about the fact that most people comply or that taxes pay for public goods like health services, roads, and schools.

We have developed a model of income tax evasion within a social network to address how the government can change its audit, penalty, and tax rates to influence perceptions and ultimately reduce tax evasion, potentially by even allowing for a temporary budget deficit. Our approach involves embedding an inductive reasoning game in an agent-based model (ABM) in which the perception that others are evading taxes is indirectly influenced by increases in taxation rates and directly influenced by social behavioral contagion mechanisms. Although ABMs have previously been applied to tax-evasion, the emphasis has been on analyzing the effects of altering audit rates and penalties. Our ABM has been designed to address a different question: what are desirable short, medium, and long-term fiscal policies that can be used together with changes in audit rates and penalties to ultimately reduce and maintain low levels of tax evasion. We model the Governing State (GS) and allow it to manage its debt in order to maintain services while attempting to transform evaders into compliant agents. We fielded a nationally-representative survey designed to directly and empirically inform our model assumptions and parameters that relate to agents’ behavior, perceptions of audit/penalty risk, tax fairness and government waste.

2 Understanding social drivers of tax evasion

Every year during tax season, individual taxpayers are faced with numerous choices regarding their tax reporting. Some of these decisions concern the ways in which income should be reported in order to minimize the tax burden, others are related to the amount of income to report and lastly, some taxpayers are even thinking about whether they should report any income at all. Depending on these decisions, some taxpayers are reporting their entire income and are complying fully with their tax burden, while others are misreporting their tax obligations and thus - either accidentally or intentionally - evading a portion or all of their tax bills. When thinking about taxpayers that are not in full compliance, one can distinguish between three scenarios:

1. Accidental misreporting describes a situation where an individual taxpayer accidentally reports his/her income in the wrong way. This type of under-reporting is found mostly when tax codes are highly complex and individual taxpayers do not rely on tax professionals to declare their taxes.

2. Tax avoidance refers to a taxpayer’s strategic efforts to minimize his/her tax burden through shifting income between various reporting categories. The most common areas of occurrence for this type of taxpayer behavior are capital gains and property or other asset-related taxes, as well as the use of tax havens and foreign revenue declarations for corporate entities.

3. While both misreporting and tax avoidance result in a failure to meet one’s appropriate tax burden and thus contribute substantially to a country’s tax gap, they are not necessarily malignant acts, but the outcomes of either legal, strategic tax behavior or oversight. Tax evasion, on the other hand, refers to a taxpaying entities’ conscious decision to misreport taxable income.
Allingham and Sandmo (1972) first developed a model to describe how individual taxpayers think of tax evasion as a gamble between themselves and the tax collection agency: if a taxpayer chooses to evade a certain percentage of his/her income, he or she faces a chance of getting caught by the tax agency, the audit rate, and if caught, will owe a certain penalty as specified in the tax law. Therefore, a tax evader minimizes expected tax burden as follows:

\[
\text{Expected(tax)} = \text{actual income} \times \text{effective tax rate} \times \left[ \% \text{ declared} + \right.
\]

\[
\left. \% \text{ not declared} \times \text{probability of being caught for tax evasion} \times \text{penalty rate} \right],
\]

whereby the probability of being caught for tax evasion depends on the audit rate, which is assumed to be increasing as the share of evaded income increases and on the detection efficiency.\footnote{The probability of being caught for tax evasion \times \text{penalty rate} represents the expected penalty rate of a noncompliant taxpayer.}

In consequence, tax evasion presents a taxpayer’s conscious effort to jointly minimize both the tax burden he or she faces and the risk of getting caught doing so. In essence, the chance that a tax evader gets caught and penalized by the tax authority can be represented as the conditional probability of receiving an audit, given that he or she is underreporting his/her tax burden. This is illustrated in figure 2.1.

There are many factors that determine how people think about taxes, such as the perceived fairness of taxes in general, their level of content with public service provision and the welfare state as a whole, as well as their perceptions of the strength of enforcement mechanisms. In particular, the probability of being audited and caught for tax evasion are commonly studied factors that affect tax compliance. Empirical studies confirm that the higher the audit rate, the higher the overall compliance \cite{8, 139, 178}. However, past work has shown that this positive relationship is not linear. These studies find that taxpayers remain compliant even when the audit rate is zero or very low \cite{9}. This is because taxpayers perceive higher risks of being audited and caught than the actual probabilities. In fact, if everyone had perfect knowledge about the actual audit rates, one would expect compliance rates to be much lower than they actually are. Instead, taxpayers seem to assume that the audit rate they face does not equal the average audit rate, but depends on their taxpaying behavior. Thus, tax avoidance and tax evasion strategies, in particular, are driven by individual taxpayers’ perceptions of the underlying tax system and the resulting risk of detection.

Taxpayer’s perceived audit rates are strongly impacted by their past audit history and that overall experiencing an audit is positively correlated with the compliance rate \cite{71, 127}. However, this is not always the case. For example, some studies have observed that taxpayers have lower compliance rates in the short-term immediately after experiencing an audit. This phenomenon is known as the bomb-crater effect \cite{86, 92, 127}. A plausible explanation of the bomb-crater effect is that taxpayers perceive a lower risk of being audited in a small window of subsequent years immediately after experiencing an audit.
Although information about actual audit rates is readily and easily available, Manhire (2013) suggests that people’s perceived audit rates are highly perspective-dependent and a direct consequence of both their own behavior as well as the behavior of others around them [113]. Thus, while the average audit rate is typically represented as the number of audits overall returns filed, the perceived audit rate that taxpayers who behave in a strategic manner consider in their calculations, should be represented as a probability that depends on both the individual’s behavior and the behavior of others around said individual. Manhire states that the two rates are only equal to each other if the IRS audit mechanism is no better than random, an idea that is inconsistent with the fact that the IRS is using a discriminant index function to maximize the share of audits leading to the detection of evaders. In consequence, strategic taxpayers assume that underreporting leads to an increase in the probability of getting caught, although they frequently have no way of knowing the difference between the two audit rates. In summary, Manhire argues that perspective-dependent audit probability theory offers a more precise measure of taxpayer risk assessments than traditional rational choice theory.

Being caught for tax evasion results in penalties. The evidence for the positive effect of the penalties on the tax compliance is less compelling than that available for audit rates. While there are some studies suggesting that the higher penalties reduce tax evasion [10, 61, 139], there are some other that found either no evidence for it or evidence supporting the opposite [4, 56, 143]. Some scholars even argue that excessive penalties may reduce tax morale, which, in its turn, negatively impact tax compliance [13, 60]. Thus, one may draw a conclusion from the literature that the deterrent effect of penalties on tax evasion is rather weak.

Another key factor is the effective tax rate which impacts tax morale and consequently impacts compliance. Higher tax rates are known to be negatively correlated with tax compliance. This effect may be even stronger if taxpayers perceive their tax rates to be unfair [129]. As with the perceived audit rate, tax morale and the perceived fairness in tax rates can be influenced by social network effects and media [7].

Along with perceptions about tax fairness, attitudes and trust toward government, social and personal norms have been receiving an increased attention as driving forces of the tax compliance in the literature. Kirchler (2007) provides an excellent account of tax studies focusing on these factors [89]. An example of more recent such studies is a paper by Shu et al. [154]. Through an experimental design, the authors suggest that making ethics salient by signing tax return at the beginning rather than at the end can increase tax compliance. Using the results of another set of experiments, Lamberton et al. demonstrated that eliciting taxpayer preferences on government spending can also increase tax compliance [96]. In another study that utilized data from a natural field experiment in the UK, Hallsworth et al. show that tax compliance can be significantly improved by simply reminding taxpayers about the fact that most people comply or that taxes pay for public goods like health services, roads and schools [74].

2.1 Shaping perceptions on tax evasion

Given that taxpayer perceptions appear to be a substantial driver of compliance and evasion choices, it is important to understand how these perceptions are generated. Furthermore, it is crucial for tax authorities such as the IRS to gain a better understanding of the various channels of information that taxpayers rely upon in developing their perceived audit rates. If the tax authority has a sound understanding of the underlying processes, it can effectively use these channels to shape the debate around tax payments in a desirable way, thereby eliciting higher rates of compliance. While an individual’s social network of friends, family, and colleagues certainly plays a substantive role in determining his/her willingness to pay taxes and the corresponding readiness
to evade [2], Branham suggests that mass media reporting can be used very effectively to achieve
tax compliance and to deter people from evading or avoiding taxes [27]. The author illustrates a
scenario where the IRS is faced with the decision to aggressively pursue one of two tax evasion
cases: the first one involves an ordinary taxpayer, while the other one involves a famous person.
Given that the case involving the famous tax evader is likely to garner a substantial amount of
public attention and media reporting, the author recommends spending resources on this particu-
lar case. In effect, pursuing cases that hold the potential for high levels of publicity can be a cheap
way of deterring future evasion. Individuals in the public eye are often highly visible and pub-
licly prosecuting them serves as both a signal of public disapproval and immorality of tax evasion
practice as well as a threat that nobody is immune from getting caught, not even rich or famous
individuals. Moreover, an experimental study by Alm et al. indicates that unofficial information
about audit experience of a taxpayer’s social network may significantly impact his/her compli-
ance [8]. More recently, Alm et al. found that taxpayers reporting decisions can largely be affected
by information they may have of the filing and reporting decisions of people they know [7].

2.2 Media’s impact on individual-level tax compliance

Kasperson and colleagues first introduce the effect of mass media on their social amplification of
risk analysis. They outlined a model to describe how media reporting and public opinion shapes
people’s assessments of various risks [85]. In the opinion of the authors, technical risk assessment
should depend only on the probability with which an event occurs, and the magnitude of the con-
sequences brought by that event. However, in reality, it can be observed that people perceive risk
in a different, non-technical way, by incorporating their own biases and subjective assessments.
In essence, the way people determine risk is a direct outcome of the interplay between cultural,
individual and social considerations. In attaining information about risky events, individuals rely
on personal experiences, media reports and their social networks to update their assessments of
event probabilities and magnitudes. Over the years, the social amplification of risk framework
has been applied to numerous areas such as public attitudes toward genetically modified foods
[59], charitable giving immediately following natural disasters [31], financial market trends [50],
brand trust in the fast-moving consumer goods industry [184], climate change [100] and the envi-
ronmental movement [1]. While all of these studies are concerned with different topic areas, they
find similar effects of media reporting. When looking at the effects of increased media coverage
for a specific issue, it appears that public attention is heightened and that people who do not have
a preconceived stance on the issue will adopt the attitudes communicated to them by their media
outlet of choice. In a sense, the media raises awareness for certain issues and provides a platform
for discussion and information exchange that is easily accessible to most.

However, while the above-referenced scholars appear to be generally in agreement about the
existence of an amplifying role of the media in shaping the perceptions of risk, Tyler and Cook
argue that media reports only affect people’s attitudes towards societal problems, but not their
personal behavior [170]. In developing their impersonal impact hypothesis, the authors use data
from three different experiments to show that media reports influence attitudes for societal, but
not for personal risks. In essence, the media serves as an indirect channel of information in the
absence of personal knowledge and experience. In all three experiments conducted as part of this
study, participants were randomly exposed to a media campaign on selected societal issues such
as crime epidemics or gun violence and were then asked a series of questions about both societal-
level and personal-level risk assessments. While on a personal level, treatment and control groups
showed virtually no differences, people’s perceptions of societal-level problems changed dramat-
ically after exposure to media reports. Tyler and Cook take this as evidence to say that while
mass media reports can be very impactful in shaping overall opinions about issues on a societal level, they are less powerful in influencing individual-level risk assessments and resulting actions. Media reports thus act as an indirect influencer when information from personal experiences and social networks are not readily available.

Nevertheless, in shaping attitudes toward tax compliance, the IRS can at the very least somewhat rely on mass media reporting to persuade people that compliance is the correct default behavior, rather than evasion. Mason uses survey data from a randomly-drawn representative sample of taxpayers in Oregon in 1975 to further underline this point [117]: he finds that while people who rely heavily on their own personal experiences tend to accurately assess their individual audit risk, individuals who rely largely on mass media as a source of information generally overestimate the chance of audit. Along those lines, Branham describes the mass media as a channel to achieve mass socialization and conditioning towards a certain norm [27]. Referencing the example of anti-smoking campaigns, she suggests that the media can play a very powerful role in determining majority attitudes and in ultimately steering behavior. In pursuing highly visible cases of tax evasion, the IRS can, therefore, very effectively, take advantage of the media multiplying factor and its ability to convey a message cheaply and quickly to a wide variety of people. Nonetheless, she also posits that a strong reliance on media reporting can be dangerous if the tax authority is not in a position to control the message. Appealing to taxpayers on normative grounds and emphasizing that paying taxes is simply "the right thing to do" and a civic duty can be a very powerful strategy to increase tax compliance as outlined by Stalans et al. [160], Kornhauser [93] and McGraw and Scholz [120], but in situations where tax evasion is starting to become the normal, acceptable behavior, the aforementioned social multiplier can work in the other direction as well, pushing more taxpayers towards evasion and away from full compliance. Accordingly, Fellner et al. [55] field a natural experiment to evaluate the effect of various different strategies to enforce compliance with the public television licensing fees in Austria, and find that appealing to morals and pointing to the high overall compliance levels in the general population are not as effective as sending threatening messages, reminding people of the substantial penalties and high detection risk.

Media coverage for tax-related issues could affect taxpayer behavior in one of two ways depending on the situation and tone of the articles: if media coverage is friendly toward the tax authority and suggests that the taxes are adequately translated into government services or emphasizes the tax authority’s high levels of efficiency in catching evaders, they can be expected to boost compliance. On the other hand, if tax-related reports center on issues of inefficiency, lack of enforcement, corruption, evasion, and waste, they could lead to lower levels of compliance. Galbiati and Zanella illustrate precisely this negative social multiplier scenario for the Italian tax system [62]. High levels of evasion in a country cause enforcement congestion and extensive media coverage reinforce the idea that the more people evade, the less likely it becomes that an individual taxpayer will be caught and penalized by the local tax authority.

While researchers have started to look into the impact of media reporting on corporate tax planning and strategies, impact on individual taxpayers has not received the same level of attention. Partially, this could be due to the fact that the directionality of the effects is not as clear: due to the fact that corporations have a public reputation and brand name to uphold and protect, they are typically pushed toward increased compliance when faced with greater media scrutiny. In consequence, researchers have started to look at the response of firms as documented in their annual reports. Notably, Chen et al. [38] sought to obtain insights into two research questions: what drives media coverage of corporate tax issues and how do firms respond in the event of negative coverage. Following a longitudinal panel of S&P 250 corporations in the United States between 2009 and 2014, the authors find that complexity of corporate tax planning efforts and opportunity
to avoid taxes as indicated by large shares of foreign sales and subsidiaries are the largest drivers of media coverage. Thus, media reporting of tax-related issues occurred largely with the purpose of informing customers, rather than providing entertainment and sensationalist stories of highly visible targets. Based on the outcomes of this first research question, the authors posit that due to the impending reputational damages from negative media coverage, firms that are subject to negative reports should react by revising their tax planning efforts. However, when carrying out the analysis, no such effect can be detected. Instead, it appears that firms might try to hide tax evasion and avoidance schemes when faced with heightened media scrutiny. In a similar study on FTSE 100 corporations between 2010 and 2013, Lee [98] exploits a natural experiment in the United Kingdom: in the aftermath of a UK Public Accounts hearing on corporate tax evaders, there was a spike in media attention on the issue. Lee uses this substantial increase in coverage to test whether companies that were at the receiving end of such reports reacted by altering their tax planning and tax communication strategies. Specifically, the volume of tax disclosures in audited financial statements and contents of tax information sections in corporate annual reports were taken as key outcome variables, standing-in as both quantity and quality measures of tax reporting practices. While once again, one would expect a strong response from firms that are in the public eye for potential tax evasion, it seems like the opposite is true, especially for firms in the financial sector. Ultimately, both studies agree that despite the potential threat to a firm’s reputation, the typical response to an increase in tax-related media scrutiny appears to be a non-response rather than an over-disclose and over-correct strategy.

Among the few studies concerned with individual taxpayer responses to media reporting, most are testing the effects of various enforcement strategies on compliance rates. As an example, Battiston et al. compare the effects of highly-publicized and quiet tax raids in various regions of Italy, using a difference-in-difference strategy [18, 19]. Essentially, the authors exploit the fact that during the first quarter of 2012, three tax raids were conducted in Milan, Turin, and Genoa (all major cities in Northern Italy), focusing on the Value-Added Tax (VAT) accounts of small to medium-sized companies. The raids in Milan were covered extensively in the media, while the raids in Turin and Genoa received little attention. As indicators of media reporting intensity, the study systematically collected data from the two major National newspapers as well as a database of all Italian newspapers, and also performed searches on both Google and YouTube. In order to assess the impact of media reporting, VAT balances (treated much like a VAT tax gap: expected - actual revenues) were compared for each region, both before and after the raids. Ultimately, the study finds that in the short term, a "loud auditing" strategy had a strong, positive impact on tax compliance, but that this effect did not persist over time. In addition, the authors describe emerging evidence of a bomb crater effect in the Genoa region, where tax compliance seemed to briefly decline sharply following the raids, before reverting back to pre-raid levels. Similarly, Kastoryano [87] is taking advantage of the fact that the Dutch tax authorities announce specific target areas for each tax season that will receive special attention in auditing practices to assess how taxpayers with a substantial share of income in those target areas respond. In particular, the study uses data from the 2005 tax season, where secondary income, freelance income not subject to third-party reporting as well as property and residual assets were announced as primary auditing targets in the Netherlands. Every year, these auditing targets are announced in large, nationwide media campaigns, about two months before the tax filing season, using a mix of newspapers, magazines, radio spots and TV announcements. In order to quantify the impact of this audit spotlight campaign, Kastoryano looks at two treatment groups and one control group: the first group is a sample of taxpayers whose primary source of income falls under the freelance work category in the Dutch tax code. The second group of taxpayers reports a substantial share of rental assets and the control group is a random sample of audited individuals. In order to construct
a longitudinal panel of taxpayers and to discern any short-term and medium-term impacts of the 2005 audit target announcements, each individual’s records were obtained for tax years 2002-2008. The main research question investigated regards the immediate response of taxpayers: given the tax audit announcement, do taxpayers divert from their prior reporting patterns and do they revert back to the old state once the 2005 tax year has passed? As one would expect, the study finds that taxpayers in the higher income brackets that have ample opportunity and incentive to declare their taxes in a strategic manner opted to substitute most of their declared income away from the tax areas in the auditing spotlight. The author takes this as preliminary evidence to say that rapid and frequent shifting between declaration categories can be understood as an indication of strategic substitution and tax avoidance patterns. However, he also posits that if taxpayers had no strategic substitutions available, the default response seemed to be increased reporting in the key tax areas.

3  Overview of ABMs and Past Models of Tax Compliance

3.1  Complex Adaptive Systems and Agent-Based Models

The system of tax evasion we described in the introduction section can be categorized as a complex adaptive system (CAS) [125]. CASs are characterized by a complex interplay between individual entities (i.e., agents), their interactions and their collective behavior through various feedback loops. They produce emergent group behavior whereby agents self-organize and adapt to new environments that they, in part, helped to form. They are adaptive because the agents have memory and they can learn from their histories. They are complex and non-linear since these new environments cannot be understood by simply summing over each agent’s behavior when considered in isolation. Moreover, non-linearity means that the magnitude of system effects need not be proportional to their causes and different effects can interfere with each other in a complex and non-additive way. CASs can be inherently stochastic as the agent’s behaviors often include chance driven effects. They are usually highly sensitive to initial conditions and parameter values and can evolve towards very different collective-level equilibrium states and exhibit critical phenomena and even chaotic dynamics.

CAS is most suitably modeled using a computational simulation that tracks individuals over time. These models provide a "bottom-up" approach where the collective-level outcomes results from aggregating the dynamics describing each individual. One category of individual-level simulation models is called microsimulation models. These models track the dynamics of individual agents over time as they make various types of transitions. The occurrence of these transitions is usually informed by transition rates that are external inputs to the model. Therefore, microsimulation models often take a descriptive approach since the transition rates are chosen with the sole aim of statistically reproducing observed collective outcome, without the need of a behavioral theory [146, 155, 158]. A different category of individual-level models is ABMs. These

---

2In contrast, a "top-down" approach, such as that used in system dynamics imposes that groups of individuals be collectively described by a set of equations and that individuals belonging to the same group act in the same way irrespective of their different histories [161]. Effects describing changes in the behavior of the individuals can be included but these are specified at the population-level and are often statistically-based and descriptive, making no claim to the real underlying behavioral mechanisms at play of the individuals. For example, in the context of an infectious disease a systems dynamics model can change a model parameter describing the level of social mixing in the population if a certain aggregated quantity, such a disease prevalence reaches a given threshold value. However, the way the parameter changes is through a high-level descriptive rule that can be informed empirically from statistical analyses of past social mixing data. This descriptive rule does not specify individual-level behavioral mechanisms such as the effects of social influences on a network.
have an important distinction in that the transitions of the agents occur due to internal casual mechanisms and behavioral rules that are endogenous to the model. An ABM approach provides the most natural computational instantiation of a CAS. This is because individuals are assigned behavioral rules that drive their interactions and interplay with the environment from which we can study the formation of individual-level heterogeneity in behavior and the emergent collective behavior. Moreover, they describe the switching of agents’ strategies and thus the lack of a settled equilibrium at the individual level.

The purpose of ABMs is to simulate the interplay between the micro and macro behavior of CASs to gain an understanding of the system which may be too complex to do using analytic methods [97]. ABM need to be verified, validated and calibrated. An empirically successful ABM will get both the micro and macro behaviors right, so the agents in the model approximate in some useful way the behaviors of agents in the real world and reproduce expected emergent properties observed under status-quo conditions. However, this is an ambitious undertaking as model validation poses a tension between getting both right. In the spirit of parsimony, it is thus important to try to keep the model as simple and as straightforward as possible [15]. Model parsimony is listed as one of the important principles of the best practices when developing ABMs [33, 75]. A calibrated ABM can be very useful for demonstrating the potential impact of policy changes and under different environments and settings, especially when the model produces revealing emergent properties and outcomes. In particular, the model can be used to explore the effects of the driving complex feedback loops and can alert us of any unintended amplifications and spill-over effects. However, the principle of model parsimony has to balance the risk of oversimplifying or neglecting fundamental internal casual mechanisms and behavioral rules. Including a larger set of fundamental behavior relationships in the model leads to a larger domain of applicability. This allows the model to address a wider range of environments and policies that can be very different than the observed current status-quo condition. As explained by Lawrence [97], ABMs should have what he calls Marschak’s stability and pass the Lucas critique [111, 116]. This essentially means that internal casual mechanisms and behavioral rules should be fundamental and general enough to allow the model to make sensible and realistic predictions under very different alternative environments and policies.

3.2 A Brief Overview of Past Models of Tax Compliance

Allingham and later Srinivasan developed the first models of income tax evasion by assuming risk-averse deductively rational individuals whose tax compliance behavior is described by a diminishing marginal utility function [5, 159]. These models assumed that individuals have perfect knowledge of audit probabilities and use this to calculate the proportion of their income they should disclose to the authorities in order to maximize their expected utility function. These first studies on tax evasion have been extended using traditional game-theoretical approaches by various authors [11, 13, 66, 67, 151, 173]. These models usually provide a static or equilibrium analysis of tax-evasion in which compliance is entirely attributed to the expected penalty of evasion,

---

3 An other approach with the same general purpose as ABM are dynamic stochastic general equilibrium (DSGE) models. DSGE models seek to find the equilibrium solutions where micro and macro level behavior reach a point of mutual self-consistency [41, 53]. Implicitly, DSGE models assumes that there exists some equilibrium towards which the system is drawn to. Further, this approach often uses other simplifying assumption. For example, they consider a small number of representative agents that use rational and deductive reasoning to correctly forecast future outcomes. Moreover, in order to best match empirical data, the self-consistent equilibrium found by using DSGE often sacrifice behavioral realism at the micro-level in favor of reproducing the macro-level behavior. Unlike DSGE, ABMs do not assume that the system can achieve a settled equilibrium or a priori assume the existence of a general equilibrium. In an ABM, no order or design is imposed by on the system from the top down.
and on the taxpayer’s risk aversion. This general approach further assumes that individuals know taxation levels and penalties, and know or can correctly deduce the audit rules and probabilities that apply not only to themselves but to everyone else in the system as well. While this might be true for taxation levels and penalties which are theoretically available to all, it is not the case for audit probabilities since tax agencies rarely disclose these. Moreover, evidence suggests that tax evasion is a dynamic process whereby agents adapt and change their tax compliance behavior based on many factors such as opportunity, social norms, perceptions of tax fairness, changes in enforcement, prior tax evasion experiences, and knowledge of tax-related experiences of other individuals. This dynamic evolution results in a population with a large variation in tax compliance behaviors, resulting from individual’s different experiences and immediate environments. In ABMs of tax-evasion individuals are assigned a behavioral rule that drives perceptions of audit probabilities and tax fairness. Building on earlier theoretical work, a few ABMs have emerged in the last decade. These ABMs can be classified into two main groups.

The first group of ABMs, designed by economists, considers a heterogeneous population with different behavioral types, different individual risk-preferences, and different taxable incomes. Agents belonging to different behavioral types (e.g., ethical/honest, imitators, free-riders, etc.) have different expected utility functions. For example, Mittone and Patelli considered agents who update their probability of compliance from year to year based upon their individual change in utility function which was modified according to their experience and the information provided by the tax agency. Inspired by the standard models of infectious diseases, Davis et al. focused on a model that includes direct agent-to-agent influences. Interestingly, they found two stable equilibria at both high and low levels of tax compliance, each occurring at different values for the audit rate. Korobow et al. consider a model in which agents have different perceived risks of being audited that are formed by observing if members within their social network are audited. They found that even low rates of enforcement may support a highly compliant society of taxpayers. Antunes et al. and Hokamp and Pickhardt both considered a model with a population with different behavioral types of which one type is a rational deductive-reasoning taxpayer. Antunes et al. concluded that social imitation, local neighborhood enforcement, and reputation are key behaviors and play a larger role than the individual perception of expected utility maximizing. Hokamp and Pickhardt explored the effects of back auditing and concluded that this has a strong effect on agents tax behavior. Bloomquist developed a detailed ABM in which agents were characterized by many attributes including income level, the degree of income visibility to the tax authority, age, memory, acquaintances, and perception of enforcement. Results suggest that a significant portion of audit-based tax evasion deterrence could come from group influences on compliance behavior. Szabo et al. modeled an emerging shadow-economy by considering an industry consisting of several competing firms and their employees. They find that the level and efficiency of tax audits alone cannot control and explain the emerging tax compliance level. Nordblom and Zamac constructed a model that considers the life cycle of agents tax behaviors. Their model supports the view that older people hold stronger moral attitudes and thus evade relatively less in their taxes. Meder et al. considered a model that focus exclusively on the effects of an exogenously set tax morale. Pellizzari and Rizzi also considered a model of tax morale where agents differ with respect to preferences for public expenditure and beliefs about the actual level of tax compliance in the population. Interestingly, they find a strong relationship between the tax rate and evasion and that individual characteristic of the agents matter more for individual compliance than audit policy parameters.

The second group of models has emerged from the area of statistical physics applied to socio-economic systems. These models coupled tax evasion behavior in the presence of an auditing
tax authority to the Ising model from statistical physics of ferromagnetism to mimic conditional cooperation among agents [188]. Zaklan et al. provided the first of these models and concluded that even minimal audit rates by a tax authority may help to alleviate tax evasion substantially and that this result holds true in spite of strong group influence [186, 187]. These models are embedded into alternative lattice and network types [104, 105]. Pickardt and Seibold constructed an Ising model of tax evasion where agents are categorized into different behavioral types [142]. These models have been extended to explore the effects of tax morale [106].

Models introduced by economists generally assume that there is a utility function that describes the behavior for each type of agent. These models assume that agents adapt and transition between behavioral types and have the parameters values that enter their utility equation change over time. However, it is possible that individuals are not consciously aware of a utility function that they wish to maximize but rather use simpler rules based on the frequency of observation of past experiences and that of their peers to make their decision. Modeling this process, which is akin to Bayesian learning approach, has the advantage that the diverse ecology of behaviors form spontaneously during the model dynamics rather than assuming the existence of different types of agents a priori. Similarly, models introduced by physicists depend on an exogenously determined variable that is equivalent to the temperature in the Ising model. This variable determines the degree to which agents make autonomous decisions. However, the interpretation of this variable remains unclear. Moreover, agents’ trust in governmental institutions was modeled by essentially applying the equivalent of an external magnetic field to the Ising system. This means that the level of tax morale was exogenously set to vary in the model and affected all agents in an equal manner. Thus, diverse perceptions of tax fairness do not emerge spontaneously in these models. Finally, a substantial empirical literature exists, including both individual-difference survey and experimental approaches, suggesting potential determinants of tax compliance. In particular, one important determinant appears to be perceived tax fairness, sometimes conceived as one component of a broader tax morale [17, 26, 121]. However, with the exception of Bloomquist’s model [24], few existing ABMs attempts to inform parameterization and other modeling decisions with primary empirical research using representative samples.

While much has been learned about tax evasion and fiscal policy using these ABMs, they do not address our specific research question. For example, none of these models explore strategies of the GS that improve compliance of taxpayers by allowing initial deficits constrained by a maximally acceptable debt. While the latter has been considered in the optimal taxation literature [2, 72, 95, 108, 112], this research does not use ABMs nor does it look at the role of tax fairness. In our model, agents’ diverse set of tax compliance behaviors will emerge endogenously via their adaptation to past experiences, observations, and interactions with peers and with interactions with the GS (including changes in tax rates). We will be addressing a very important and timely question in fiscal policy using for the first time a behavioral ABM approach.
Overview, Design concepts, and Details

4 Purpose

A taxpayer’s propensity to engage in income tax evasion is determined by perceptions of deterrence (i.e., risks of being audited & penalized) as well as the perception of the fairness of the tax system. These perceptions change over time leading to changes in behavior and attitudes towards tax compliance. Such perceptions and their formation depend on (i) personal evaluations of tax compliance (e.g., tax morale, tax rate, audits and penalties), (ii) social interactions, including observations and influences by alters in one’s social network (SN), and (iii) media feedback of the severity of tax evasion and the efficiency in catching evaders. Understanding the interplay between perceptions formation, personal evaluations, social interactions and media feedback is critical to informing effective fiscal and deterrence policies to increase tax compliance.

We constructed an agent-based model (ABM) of income tax evasion within a social network to address how government can change its audit, penalty, and tax rates as well as other policies to influence perceptions and ultimately reduce tax evasion, potentially by even allowing for a temporary budget deficit. Our aim is to first use the model to answer a set of questions that could help to better understand tax compliance dynamics. These questions include (a) what are the determining assumptions and parameter values describing agents behavior that have the most leverage on the compliance levels, (b) what is the role of social networks in determining the compliance level that the system will equilibrate towards, and (c) what are the conditions for critical behavior in our system and which parameters determine critical points. We linked the model describing the dynamics of how taxpayers change their compliance behavior to a model describing government policy.

Our aim is to use the model to find socially acceptable and economically desirable strategies that can transform a population of taxpayers from one with a majority of evaders to one with a majority (or significant increase) of compliant agents. Specifically, we are interested in using our model to answer or gain insights into the following question: Starting from zero debt in a society characterized by significant tax evasion, (a) can tax evasion be reduced by a specified amount within a given time frame and (b) if so, what policy minimizes the accrued debt? (c) what is the shortest time required to repay the debt? The time to repay the debt, and the debt itself represent the regret measures of the policy intervention which need to be minimized within the constraints of being an acceptable strategy. Achieving this aim is not trivial: strategies that aim to maximize short-term rewards may not necessarily be desirable in the long term. In a society in which tax evasion is rampant, a short-term strategy might increase audit rates and penalty rates while decreasing taxation levels. As illustrated in Figure 4.1, this strategy would increase the national debt in the short run but might increase compliance in the long run. Eventually, taxation rates may have to increase to repay the debt, and if done at a slow enough rate, this strategy could ensure that compliant taxpayers were not lost in the process.

A desirable policy combines the concept of optimality with robustness. We use methods and
techniques taken from the Robust Decision Making (RDM) framework to rank policies and describe their trade-offs based on both their performance and how robust they are to the uncertainties and heterogeneities of some of our behavioral parameters and network structures. RDM focuses on situations in which the ultimate goal of the computational model is not predictive simulations, but rather informing better decisions in situations in which deep uncertainties make predictions unreliable. RDM is particularly well suited for analyzing models of complex systems [83, 101, 102]. For our purpose, the RDM framework helps in identifying the most reliable strategies which perform well and are least sensitive to the most significant uncertainties.

An initial key step of an RDM analysis is to organize and structure the decision and the factors considered in the model development, exploration, and its analysis. This is done by an "XLRM" table whose components are defined as follows:

- Exogenous uncertainties (Xs) are factors outside the control of decision-makers that may nonetheless prove important in determining the success of their strategies.
- Policy levers (Ls) are near-term actions that, in various combinations, comprise the alternative strategies decision-makers want to explore.
- Relationships (Rs) are potential ways in which the future, and in particular those attributes addressed by the measures, evolve over time based on the decision-makers choices of levers and the manifestation of the uncertainties. A particular choice of Rs and Xs represents the future state of the world.

<table>
<thead>
<tr>
<th>Exogenous Uncertainties (X)</th>
<th>Policy Levers (L)</th>
</tr>
</thead>
<tbody>
<tr>
<td>The assumption describing the compliance behavior of taxpayers. For example:</td>
<td>Government and IRS policies, and other &quot;what-if&quot; scenarios. For example:</td>
</tr>
<tr>
<td>• Taxpayer’s perceived fairness of the tax rate.</td>
<td>• Deterrence policy: Audit and penalty rates, the audit strategy.</td>
</tr>
<tr>
<td>• Taxpayer’s social network structure.</td>
<td>• Fiscal policy: marginal tax rates and tax brackets.</td>
</tr>
<tr>
<td>• Taxpayer’s perceived risks of being audited and penalized.</td>
<td>• Tax amnesty policies.</td>
</tr>
<tr>
<td>• How the IRS selects taxpayers to be audited.</td>
<td>• Taxpayer’s Income distribution and proportion of self-employed filers.</td>
</tr>
<tr>
<td>• How the Media reports tax related aggregated outcomes.</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Relationships (R)</th>
<th>Measures (M)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relationships and feedback loops between the entities that enter the model and affect taxpayer’s compliance behaviors. For example:</td>
<td>Main model outcomes include:</td>
</tr>
<tr>
<td>• Effects of audits and penalties.</td>
<td>• Tax gap</td>
</tr>
<tr>
<td>• Social network contagion effects.</td>
<td>• The distribution of the proportion of hideable income reported by taxpayers.</td>
</tr>
<tr>
<td>• Effects of Media reports.</td>
<td>• Persistence of taxpayer’s compliance behaviors.</td>
</tr>
<tr>
<td>• Past experiences with tax compliance and the IRS.</td>
<td>• Population-level average perceived audit rates.</td>
</tr>
<tr>
<td>• How the GS and the IRS change their policies based on aggregated outcomes.</td>
<td>• GS and IRS outputs including tax revenues, audit costs, deficit and the debt.</td>
</tr>
</tbody>
</table>

Table 4.1: XLRM table describing the factors explored using the model.
• Measures (Ms) are the performance standards that decision-makers and other interested communities would use to rank the desirability of various scenarios.

Table 4.1 illustrates our XLRM table and lists the main entries for each component. The entries in each list will be described in detail in different parts of the report. For example, assumptions and parameters of the taxpayer’s behavioral model that enter the X-component are described in detail in sections 6, 12 and 11.2. R-components are described in sections 5 and 6. L-components are described Sections 21 and 22, and M-component are described in sections 9 and 18.

5 Entities, State variables, and Scales

Our model considers four entities (i.e. agent types), namely (i) the taxpayers, (ii) the tax-collecting agency, (iii) the government or governing state (GS) and (iv) Media. Our model focuses on the U.S. settings. Therefore, we refer to the tax collecting agency as the Internal Revenue Service (IRS).

5.1 The Taxpayers

Taxpayers represent our main agents. We also refer to them as individuals or simply by agents. Our model considers a heterogeneous population made of N taxpayers that every year are required to report and pay federal income taxes. Taxpayers may have dependents. The total population of taxpayers and their dependents is denoted by M. The taxpayer population is diverse across various static attributes which include income, family structure, employment type and proportion of hideable income. Each taxpayer i has an adjusted gross income (AGI) \( I(i) \) and is required to pay a fraction of his/her income in taxes to the IRS every year. This fraction is denoted by \( T(i) \) and is the effective tax rate applied to taxpayer i and depends on the filing status and on the marginal tax rates that the GS sets for each income tax bracket and for each filing status. We consider each taxpayer to be in one of four filing statuses: single, head of the household, married filing separately and married filing jointly. Since the GS can change the income tax bracket and the marginal tax rates, the effective tax rate can depend on time and this dependence is denoted by the subscript \( t \). However, the actual effective tax rate \( T(i)_t \) applied to the taxpayer depends on the income s/he reports to the IRS rather than on his/her true AGI and this is considered by our model. The reported income of taxpayer \( i \) in year \( t \) is denoted by \( A(i)_t \).

An agents’ employment type can be either self-employment or non-self-employed (i.e., work for a wage).\(^5\) The self-employed may choose to underreport their entire income. Therefore, the proportion of the hideable income of the self-employed is 100%. Those who are non-self-employed are assumed to have an employee that pays them a salary which is assumed to be known by the IRS via third-party reporting. However, they may still have other sources of income that they could hide from the IRS (e.g., income from the rent of a property). The proportion of hideable income for those who work for a salary depends on the income.\(^6\)

Taxpayers interact in a virtual space. They are connected in a nondirectional social network whereby they can mutually affect each other’s tax-related views and perceptions. Each taxpayer \( i \) has a fixed static set of tax-related social ties or alters denoted by \( J(i) \) on the social network. Taxpayer \( j \) is an alter of \( i \) if \( j \in J(i) \) and consequently \( i \in J(j) \). Alters represent specific social

---

\(^4\)A fifth type would be certified tax preparers which we currently do not have in our model.

\(^5\)In the US, the self-employed taxpayers are those that fill out an IRS Schedule C headlined "Profit or Loss From Business (Sole Proprietorship)" to report how much money they made or lost in their business.

\(^6\)The way we assign hideable income to taxpayers that are salaried will be described in Section 11.1.
contacts of the taxpayer and include those taxpayer s/he is comfortable and at ease with in terms of talking about all matters that pertain to taxes. To account for assortative mixing, the attributes for employment type and income determine how our taxpayers are connected in the network.

Taxpayers also have additional static and dynamic attributes. These generally describe (i) their risk perceptions of being audited and penalized by the tax-collecting agency and (ii) their tax fairness and morale. The latter include taxpayer’s attitudes towards the taxes they pay, whether they received a tax refund, past experiences with the IRS on being penalized for unreported income, the public services they receive for the taxes they pay and their perception that other taxpayers are paying their fair share in taxes. We will list and explain all these attributes in the Section 6 that describes the taxpayer’s behavioral model. The dynamic version of these attributes includes the perceived audit and penalty rates and the propensity to comply. Each taxpayer i has his/her own perceived audit and penalty rate for year t, respectively denoted by \( \tilde{q}_t(i) \) and \( \tilde{P}_t(i) \). The perceived deterrence level in year t of taxpayer i is denoted by \( \tilde{x}_t(i) \) and is equal to the product of perceived risk of being audited and penalized if caught evading (i.e., \( \tilde{x}_t(i) = \tilde{q}_t(i) \tilde{P}_t(i) \)). The static version of these attributes represents the taxpayers baseline initial values.

A taxpayer’s perceived deterrence and his/her tax morale affect his/her future compliance behavior. Taxpayer i’s propensity to comply in year t is denoted by \( w_t(i) \) and for compliant agents, it is used as a probability to initiate in under-reporting. Taxpayers that are caught by the IRS and are penalized for tax evasion change both their perceived deterrence and their tax morale. These effects separately contribute to an increase in \( w_t(i) \). The effect on tax morale is based on the amount of past unpaid taxes due and the penalty to pay.

Our model does not account for vital dynamics. Therefore, we do not explicitly model the death and replacement of taxpayers. Instead, in order to account for the key effects of vital dynamics we consider a model parameter that describes the tax-paying generation lifetime, or rather its half-life. This is the longest time scale in our model and is several decades long. Over this timescale, all experiences of being audited and penalized and the acquired information by taxpayers such as of the true penalty rate are lost. Furthermore, the model includes an option that allows for a new set of alters to be stochastically reassigned to each taxpayer after a tax generation lifetime. Thus, for all intents and purposes, the taxpayer is effectively replaced by a new taxpayer with a similar set of static attributes. Our model considers shorter time-scales where the taxpayer’s perceptions of tax fairness and risks of being audited decay to baseline values in less time, typically over several years.

5.2 The Internal Revenue Service (IRS)

The IRS is responsible for collecting taxes and enforcing the tax code. In our model, the aim of the IRS is that (i) compliant taxpayers remain compliant and (ii) noncompliant taxpayers get caught, get penalized and become compliant. To achieve their goal they rely on a deterrence policy. As part of their policy, every year they select taxpayers to audit based on a given criterion which we call the audit strategy. These criteria could depend on various taxpayers attributes such as their income and their employment type. An IRS audit can either be via correspondence or via a field audit. Correspondence audits are usually about 75% of all audits and only check for compliance in the previous tax year. Instead, field audits may include various years of back-audits. Therefore, taxpayers that are selected for a field audit will have their latest tax records and those from the previous K years investigated by the IRS. Taxpayers that are found to have significantly under-reported their income are considered tax evaders and are penalized by the IRS. In addition to the audit strategy, the IRS can also change the number of years of back-auditing and the penalties.
The proportion of taxpayers that are selected every year for an audit is called the audit rate and this depends on how much funding the IRS receives every year from the GS.

The IRS tracks overall compliance by measuring the yearly gross tax gap which is the amount of true tax liability that is not paid voluntarily and timely. The gross tax gap excludes penalties and interest. Expressed as a proportion, the tax gap is the ratio between the net underreported tax and the total true tax liability (i.e., the sum of the net underreported tax and the tax reported). In our model, the IRS also tracks a total tax gap measure equal to the gross tax gap plus the taxes that are collected as the result of its enforcement activities including penalties and interest. We denote the gross tax gap in year $t$ as $\Xi_{t}^{\text{[gross]}}$ and the total tax gap as $\Xi_{t}^{\text{[tot]}}$.

5.3 The Governing State (GS)

The aim of the Governing State (GS) is to cover the costs of providing services to taxpayers while minimizing the tax evasion, yearly deficits, and national debt. We assume that the GS provides and maintains a fixed set of services (i.e., schools, roads, policing, etc.) that have a total yearly fixed cost $C^{\text{[Fixed]}}$. The GS can change the yearly funds it provides the IRS to carry out its auditing activity and thus incurs a variable cost $C^{\text{[Variable]}}$. The GS can also change its fiscal policy by changing the tax brackets and the tax rates applied to each tax bracket. Changes in fiscal policy will have an effect on both taxpayer’s tax compliance behavior and the GS costs and projected revenues.

5.4 The Media

The aim of media is to report news to the taxpayers by informing them about (i) aggregated population levels of tax evasion and (ii) extraordinary and unusual activities levels of the IRS audit investigations and catching tax evaders. The media monitors the gross tax gap and if it exceeds a certain threshold level it begins to inform taxpayers about the problem of tax evasion with an emphasis or intensity that depends on the tax gap itself. This modifies the taxpayer’s tax morale as they increasingly become aware of the national tax evasion problem and that others in the population are not paying their fair share in taxes.

The media also monitors two other indicators: (i) the yearly number of audits and (ii) the aggregated amount of penalties and recovered taxes from evading taxpayers. If in the current year either value of these two indicators is unusually high, the media will inform the taxpayers and will report how the unusual level of audits and penalties are distributed across income brackets. In our model, most taxpayers only pay attention to these media reports if a significant proportion of the number of audits involved taxpayers belonging to their same income bracket. Others instead always pay attention to these media reports regardless of income bracket. Thus, these taxpayers will always be affected by these media reports even if they belonged to the lowest income bracket and even if the media reports that those who were audited and penalized all belong to the highest income bracket. These taxpayers are affected by what we refer to as the famous person/actor effect and are identified in our model by a static attribute called the actor attribute. Taxpayers that pay attention to these media reports will respond by increasing their perceived risk of being audited and may become aware of the true penalty rate.

\footnote{The total tax gap measure is similar to the net tax gap but it includes penalties and interest.}
6 Model Overview and Scheduling

6.1 General Overview

We begin by describing the general overview of the model. For ease of access, the number labels describing the separate steps in our model are also included as comments in the main file of our code. Figure 6.1 shows a schematic overview of the model highlighting the major feedback loops. The model implementation proceeds iteratively from year to year, which we respectively denote by years $t$ and $t+1$.

![Figure 6.1: Model overview.](image)

Figure 6.1: Model overview. Taxpayer’s outcomes depend on their compliance decisions and whether or not they are audited and caught by the IRS for tax evasion. The IRS deterrence strategy is characterized by the audit rate $q$, how they select taxpayers to be audited (i.e., the audit strategy) and the penalty rate $P$. Individual outcomes influence the taxpayer’s tax evaluations, changing their tax morale and their perceived risk of being audited and penalized in the future. Taxpayers tax evaluations are shared with their alters on their social network with which they are at ease to talk about tax-related issues and outcomes. Their experience and evaluation modify their propensity to comply in the next year. The overall aggregated effect of taxpayer’s compliance behavior determines the tax gap which can also affect the taxpayer’s tax-related evaluations via changes in government fiscal policies and by the media reporting back about the level of the problem of tax evasion in the population.

**Step 1: The tax compliance decision process.** The model implementation describing the iterative process starts with the taxpayer’s behavioral model (TBM) which proceeds by an ordered process of evaluations, decisions, and outcomes. For ease of presentation, we will explain the TBM for iteration $t + 1$ in Section 6.2. Here we just mention that at the end of the decision process, each agent will have decided whether to report all of their income to the IRS or under-report their income and if so, by how much. Their decision is based on a quantity that we call the propensity to comply that is updated by the TBM. Agent’s $i$ propensity to comply is denoted by $w_i^{(t)}$. Compliant and non-compliant taxpayers update and use their $w_i^{(t)}$ in different ways. Compliant taxpayers
first use \(w^{(i)}_t\) as a probability to initiate in under-reporting and thus tax evasion. Non-compliant taxpayers, including those who just decided to evade use their \(w^{(i)}_t\) to determine the intensity of their evasion. The main outcome of this decision is the amount of income \(A^{(i)}_t\) that they chose to report to the IRS and thus the amount of taxes they pay on the reported income. We denote \(a^{(i)}_t\) as the amount of income that taxpayer \(i\) reports in year \(t\) expressed as a proportion of his/her hideable income.

**Step 2: The activities of the IRS.** As shown in figure 6.2a, taxpayer’s outcomes are affected by the audit activities of the IRS. The IRS selects taxpayers to be audited based on the audit rate \(q_t\) and their audit strategy. They check the amount of income reported by the selected taxpayers in the current year for correspondence audits and for the current and past \(K\) years for field audits. Audited non-compliant taxpayers are caught by the IRS with a probability that depends on the average detection efficiency. Our model considers both correspondence and field audits which have different detection efficiencies described by parameters \(\epsilon^{\text{corr}}_A\) and \(\epsilon^{\text{field}}_A\). If caught evading, a penalty represented by a fraction \(P_t\) of each year’s tax underpayment is added to the total taxes due. Thus, \(P_t\) is also referred to as a penalty rate. We assume that taxpayers pay all their back taxes and penalties in the current year \(t\) on any detected unreported taxable income. Penalized agents evaluate a quantity \(\Delta^{(i)}_{\text{Penalty}}\) which is equal to the total past taxes due to the IRS plus the penalties expressed as a proportion of the yearly taxes due if the taxpayer had been compliant. The value of \(\Delta^{(i)}_{\text{Penalty}}\) provides a relative measure of how much the tax penalty affected taxpayer \(i\) and had an impact on his/her the personal finances. Section 7.2 describes the calculation of \(\Delta^{(i)}_{\text{Penalty}}\). At the end of the IRS activities, individual outcomes will include whether an agent was audited and whether they were penalized.

In addition to audit and penalty activities, the IRS provides taxpayers with tax refunds. In our model, taxpayers can receive a tax refund from the IRS when they file their income taxes. Other taxpayers may instead have a balance due and be required to pay. We assume a simple Markov process, whereby taxpayers can stochastically transition from year to year from a refund return to a balance due. Taxpayers that had a refund return the previous year have a probability \(\rho_{r\rightarrow a}\) that they will not receive a refund return this year, and those instead that did not receive a refund return and were required to pay a due balance have a probability \(\rho_{a\rightarrow r}\) that they will receive a refund return in the present year. The other two corresponding probabilities are calculated as \(\rho_{r\rightarrow r} = 1 - \rho_{r\rightarrow a}\) and \(\rho_{a\rightarrow a} = 1 - \rho_{a\rightarrow r}\). These transition rates depend on the taxpayer’s income bracket. Our model does not track the amount of the tax refund.

**Step 3: Update perceptions.** The way that agents update their risk perceptions is based on whether they (i) experienced being audited and possibly penalized, (ii) know of anyone in their social network that was audited and (iii) were exposed and responded to any news that the media releases on IRS auditing and penalty activity. We provide a detailed account of how the risk perceptions are updated and the role of the media in Sections 7.5 and 7.6. Agent’s outcomes which include their updated risk perceptions provide the first feedback loop that affects the taxpayer’s compliance behaviors in year \(t + 1\).

**Step 4:** By aggregating the tax revenues the model calculates the tax gap and the budget deficit and updates the total debt of the GS. As illustrated in figure 6.1, the GS, and the IRS can react to these aggregate measures and revise their fiscal policy and deterrence strategy. The GS can change (i) the tax brackets, (ii) the marginal tax rates applied to each tax bracket and for each filing status, and (iii) its funding level to the IRS. Consequently, depending on the funding level received, the IRS can change (i) the overall audit rate for the next year \(q_{t+1}\), (ii) the deterrence strategy by changing the way they select the taxpayers that get audited (i.e., the audit strategy),

24
and (iii) the penalty rate $P_{t+1}$ for the following year. Changes in the fiscal policy and deterrence strategy provide a second feedback loop to the TBM and taxpayer’s compliance behaviors in year $t + 1$.

**Step 5:** The third feedback loop to the taxpayer’s compliance behaviors comes from the media. As mentioned previously, there are two sets of indicators that the media monitors. This is illustrated in figure 6.3. The first is the gross tax gap. If the tax gap $\Xi_{t}^{\text{gross}}$ exceeds a certain threshold level $\Theta_M$ it informs taxpayers about the problem of tax evasion. Consequently, taxpayers may increasingly become aware that other taxpayers in the population are not paying their fair share in taxes. An indicator variable $I_M$ is set to 1 when this threshold level is exceeded and is set to 0 otherwise. This indicator variable is used to flag whether the media is active and informing taxpayers about the problem of tax evasion in the population. Our model implementation calculates a variable denoted by $\Delta_M$ that ranges in $[0, 1]$ that expresses the media relevant taxpayers evaluation to fully comply based on the knowledge of the current gross tax gap. When the tax gap is large, this media feedback effect results in a low value of $\Delta_M$ that is close to 0. Section 7.3 explains how $\Delta_M$ is calculated and its dependence on the gross tax gap $\Xi_t^{\text{gross}}$.

The second effect comes from the media monitoring the audit rate and the amount of recovered taxes and penalties assigned by the IRS every year. Using a simple $z$-test, when either of these
two indicators becomes statistically larger than their moving average over a large period of time denoted by $\tau^{\text{(Media)}}$, the media reports to the taxpayers the news of the extraordinary activities levels by the IRS in auditing or catching tax evaders, potentially modifying taxpayers perceived risks of being audited by the IRS in the next year. However, when the media reports this news most taxpayers will not modify their perceived risk of being audited $\tilde{q}^{(i)}_t$ unless they are also informed that a significant proportion of taxpayers that were audited belonged to their same income bracket or to lower income brackets. However, a minority of taxpayers are affected by the media and modify their perceived risk of being audited irrespective of the breakdown by income bracket. We assume that these taxpayers are affected by what we call the famous person/actor effect: if they learn that the amount of recovered unpaid taxes and penalties is unusually large they modify their perceived risk of being audited even though the IRS audits and penalties are targeting tax evaders that have much larger incomes. Taxpayers affected by the famous person/actor effect are roughly 15.5% of the population and are identified using a regression model of our survey data described starting from Section 13. The details of how the perceived risk of being audited is modified by media are explained in Section 7.4 and 7.5.

**Step 6:** The final step computes and updates outputs. We describe the main model outputs in Section 18. Most model outputs do not affect the evolution of the model and are used to provide summary statistics and individual-level trajectories. However, our model is designed in a way that it has an option which allows the GS to react to some of the aggregate-level model outputs such as the tax gap and the debt. The GS can then change its fiscal deterrence policy. This is illustrated in Figure 6.2.
6.2 The Overview of the Taxpayer’s Behavioral Model (TBM)

In this section we describe the TBM in step 1 for iteration year $t + 1$. At the beginning of year $t + 1$ (and subsequent time steps) taxpayers go through the evaluation and decision process illustrated in figure 6.4 which represents the main part of the TBM.

![Figure 6.4: Flow diagram showing the main components of the taxpayer’s behavioral model.](image)

**Step 1.1**: Taxpayers first make tax and risk considerations. They make a personal level evaluation by considering (i) their effective tax rate, (ii) whether they evaded paying some or all their taxes in the previous year, (iii) their perceived audit and penalty rates $\tilde{q}^{(i)}_t$ and $\tilde{P}^{(i)}_t$, and (iv) whether they received a tax refund on their tax returns. The latter is only considered by taxpayers who filed a non-zero amount of income in their previous year’s tax return. All these components are used to find a taxpayer’s personal evaluation which is denoted by $\Delta^{(i)}_p$ and its value ranges in $[0, 1]$. Taxpayer $i$ is more likely to comply if their $\Delta^{(i)}_p$ is large and close to 1. The way that $\Delta^{(i)}_p$ is calculated from each of these four components will be described in Section 7.1.

**Step 1.2**: Taxpayers interact socially with their alters regarding their tax morale as described by their value for $\Delta^{(i)}_N$. This interaction is implemented in the model by the network level evaluation denoted by $\Delta^{(i)}_N$. It is calculated by taking the average value of the personal evaluations $\Delta^{(j)}_p$ of
taxpayer i’s alters’ (i.e., \( j \in J^{(i)} \)). This can be mathematically expressed as

\[
\Delta_N^{(i)} = \langle \Delta_P^{(j)} \rangle_{j \in J^{(i)}} = \sum_{j \in J^{(i)}} \Delta_P^{(j)} / \mathcal{M}(J^{(i)}),
\]

where \( \mathcal{M}(J^{(i)}) \) is a normalizing factor giving the number of alters that taxpayer \( i \) interacts with about taxes. We also calculate an indicator variable \( I_N^{(i)} \) with is equal to 1 if taxpayer \( i \) does have at least one alter to interact with and is equal to 0 otherwise.

**Step 1.3:** Taxpayers consider their personal evaluations \( \Delta_P^{(i)} \), their network evaluations \( \Delta_N^{(i)} \) and media feedback \( \Delta_M \) and weigh each of these three components respectively by \( \beta_P, \beta_N I_N^{(i)} \) and \( \beta_M I_M \). The weighted sum of these three components provides an overall present evaluation for compliant taxpayers and taxpayers who haven’t been caught and penalized by the IRS. The overall present evaluation for a taxpayer \( i \) who is either compliant or has not been caught and penalized by the IRS can be expressed as

\[
\Delta_t^{(i)} = \frac{\beta_P \Delta_P^{(i)} + \beta_N I_N^{(i)} \Delta_N^{(i)} + \beta_M I_M \Delta_M}{\beta_P + \beta_N I_N^{(i)} + \beta_M I_M}.
\]

The parameters \( \beta_P, \beta_N \) and \( \beta_M \) are weights that add to 1 and provide the relative importance that taxpayers place on each of these three experiences, interactions or sources of information that affect their tax morale. However, since not all our taxpayers have network interactions, and since media does not report about the tax gap and about the evasion rate in the population unless these reach high levels, the true network and media weights are \( \beta_N I_N^{(i)} \) and \( \beta_M I_M \). Consequently, the weighted sum is normalized by \( \beta_P + \beta_N I_N^{(i)} + \beta_M I_M \).

Taxpayers that are caught evading and are penalized by the IRS do not use the expression shown in equation 6.2 for their overall present evaluation. Instead, their \( \Delta_t^{(i)} \) is set equal to a potentially much larger value which depends on the amount of past unpaid taxes due to the IRS and on the penalty itself. Being caught and penalized for tax evasion is a very salient experience that strongly diminishes the relative importance of experiences and evaluations that occurred in the years before being penalized. Thus, in our model when a taxpayer is penalized his/her \( \Delta_t^{(i)} \) is set equal to \( \phi \Delta_{\text{penalty}}^{(i)} \). Section 7.2 describes how the value of \( \Delta_{\text{penalty}}^{(i)} \) is calculated and the role of the scaling parameter \( \phi \).

**Step 1.4:** The value \( \Delta_t^{(i)} \) represents taxpayer i’s present overall evaluation due to outcomes and consideration from year \( t \). Taxpayers remember and weight previous evaluations with respect to present evaluation. Therefore, \( \Delta_t^{(i)} \) is used to update taxpayer i’s propensity to comply \( w_t^{(i)} \) which is determined by an exponentially weighted moving average (EWMA) of past evaluations. This weighted sum represents taxpayer i’s pro-compliance experience \( V_t^{(i)} \), which is defined and updated according to

\[
V_{t+1}^{(i)} = sV_t^{(i)} + \Delta_t^{(i)} = \sum_{m=1}^{t} s^{t-m} \Delta_m^{(i)}.
\]

The parameter \( s \) which ranges in \([0, 1]\) discounts the previous year’s evaluation with respect to the evaluation of the present year. When \( s \) is equal to 0, taxpayers completely ignore the outcome of

---

8In our current version of our model, these weights are the same for all taxpayers and therefore are not a taxpayer’s attribute.
previous years. When \( s \) is equal to 1, taxpayers give equal weight to the outcomes from previous years as the present outcome \([152]\). The discount parameter can be expressed in terms of a half-life \( \tau \) such that \( s^\tau = 0.5 \) and hence \( s = e^{-\log(2)/\tau} \). For example, if \( s = e^{-\log(2)/2} \) (i.e., with a half-life of 2 years) past overall evaluations of three years ago are half as influential on the decision-making as the overall evaluation of the previous year. The pro-compliance experience is used to update the compliance probability \( w_t^{(i)} \) for year \( t+1 \) as follows

\[
w_{t+1}^{(i)} = V_{t+1}^{(i)} / \mathcal{N}_{t+1}(s).
\] (6.4)

The term \( \mathcal{N}_{t+1}(s) \) is simply a normalizing factor representing the maximum value \( V_{t+1}^{(j)} \) could be in year \( t+1 \). As long as the values for \( \Delta_m^{(i)} \) are smaller than 1, the normalization is given by the simple sum of a geometric progression \( (1 - s^{t+1}) / (1 - s) \) which in the long time limit is equal to \( 1 / (1 - s) \sim \tau / \log(2) \). Appendix A provides more details of this general approach and explains why this is an EWMA.

Compliant and non-compliant taxpayers update and use their \( w_t^{(i)} \) in different ways. As will be described in Section 7.1, this is because compliant and non-compliant taxpayers are likely to have different overall evaluation values for \( \Delta_t^{(i)} \). Compliant taxpayers, including taxpayers that were caught and penalized for tax evasion in year \( t \) use \( w_t^{(i)} \) as a probability to determine whether to initiate in tax evasion. Non-compliant taxpayers, including those who just initiated in tax evasion use their \( w_t^{(i)} \) to determine the intensity of their evasion and thus the amount of income that they chose to under-report. Thus, the proportion of hideable income \( a_t^{(i)} \) that taxpayer \( i \) choses to report in year \( t+1 \) is

\[
a_{t+1}^{(i)} = \begin{cases} 1 & \text{with probability } w_{t+1}^{(i)} \text{ if taxpayer } i \text{ was compliant in year } t, \\ w_{t+1}^{(i)} & \text{if taxpayer } i \text{ was not compliant in year } t \text{ or is initiating in tax evasion.} \end{cases}
\] (6.5)

Consider compliant taxpayers that are on the brink of initiating or relapsing back to tax evasion behavior. For taxpayer \( i \) belonging to this set, the propensity to do so is equal to \( 1 - w_{t+1}^{(i)} \). Equation 6.5 implies that for those who do decide to initiate in tax evasion, the intensity of their evasion is equal to what was used as their propensity to initiate in tax evasion that year. Here, by intensity we mean the amount they underreport as a proportion of their hideable income (i.e., \( 1 - a_{t+1} \)). In subsequent years, taxpayers who do evade on their taxes continue to do so unless their value for \( w_{t+1}^{(i)} \) returns to being equal to 1 or are caught and penalized for tax evasion.

### 6.3 The Model Code

The model was developed entirely in R. The code is available externally on a public git server. Details on accessing the code and the file structure are provided in Appendix K. The repository also contains an R Shiny interactive application that was developed for the purpose of model verification and to provide quick visualizations of the model results starting from a calibrated stationary condition. A description of the tool is provided in Appendix L.

29
7 Specifications of the Taxpayer’s Behavioral Model Components

7.1 The calculation for the taxpayer’s personal evaluation $\Delta_P^{(i)}$

Fairness concerns are the most frequently mentioned topics when citizens are asked what they think about the tax system [89]. Therefore, fairness considerations represent the first step in a taxpayer’s personal-level evaluation. Perception of tax fairness may decrease if a taxpayer perceives (i) a higher tax rate, (ii) a lower level of public services that they wish to consume, (iii) tax complexity, bureaucracy and corruption in the system and (iv) that other taxpayers avoid paying their fair share in taxes. Our model focuses on the taxpayer’s perception of the tax rate and on their perception that other taxpayers are compliant. Social influences and media effects influence their perception that other taxpayers are compliant. These enter the model via the network and media level evaluations $\Delta_N^{(i)}$ and $\Delta_M^{(i)}$ respectively. A taxpayer’s personal-level evaluation $\Delta_P^{(i)}$ is instead affected by the perceptions of his/her effective tax rate and whether or not s/he received a tax refund from the IRS and enforcement parameters. Here we explain how $\Delta_P^{(i)}$ is evaluated starting from the perception of the fairness in the effective tax rate.

Each taxpayer in our model has a different opinion of what is the fair effective tax rate that should apply to them. We assume that in the absence of any perceived deterrence each taxpayer $i$ has a maximum effective tax rate $c_1^{(i)}$ above which s/he would consider taxes to be unfair. Beyond the threshold value $c_1^{(i)}$, perceived fairness is assumed to fall monotonically with increasing effective tax rate. Eventually, perceived fairness reaches a value of zero when the effective tax rate reaches a second threshold value $c_2$, and it will remain at zero for larger effective tax rates. We assume that this second threshold value $c_2$ is the same for all taxpayers and represents the point when taxes are too high and exasperating for all taxpayers to either continue to work for pay or to continue to report their income to the IRS and pay their taxes. For simplicity, in our current model, we assume that fairness decreases linearly between the two thresholds $c_1^{(i)}$ and $c_2$. However, we could equally use an S-Curve function as we have considered in our analysis of the survey data described in Section 14.

This effective tax threshold value $c_1^{(i)}$ may well depend on what each taxpayer perceives to receive as public services in return and what instead is wasted through tax complexity, bureaucracy, and corruption in the system. However, our model does not consider cases where the levels of public services received, tax complexity, bureaucracy and corruption change over time and consequently we assume that their perceptions of government waste and bureaucracy also remain constant over time. Figure 7.1 illustrates how the function describing how perceived fairness changes with effective tax rate looks for a taxpayer with $c_1^{(i)} = 15\%$ (black solid line) and how the value of $c_1^{(i)}$ may be distributed in the population (blue histogram). We refer to this function as the fairness function and note that it can only be called a fairness function when the taxpayer’s perceived deterrence rate is zero.

In presence of deterrence, taxpayers perceive risks and consequences of under-reporting their income to the IRS. Consequently, they are more willing to fully report their income and would only start to consider under-reporting when their effective tax rate reaches a larger threshold value. We denote this new threshold value as $\tilde{c}_1(\tilde{x}_i^{(i)})$ or simply by $\tilde{c}_1^{(i)}$. It is bounded by $[c_1^{(i)}, c_2]$ and it

\[9\]

Since we are considering the hypothetical case of no perceived deterrence, we could think of $c_1^{(i)}$ as a donation rate rather than a tax rate. Thus, $c_1^{(i)}$ is the maximum proportion of the income of taxpayer $i$ that s/he is willing to donate to the government to support its activities and in providing public services that they s/he and other citizens consume.
Figure 7.1: The step-wise linear black lines show the function $F(T_t^{(i)}, \tilde{x}_t^{(i)})$ for one taxpayer and for the cases where s/he perceives no deterrence (a) and where s/he perceives a positive non zero deterrence (b). In this example the threshold value moves from a value of $c_1^{(i)} = 15\%$ in (a) to a value of $\tilde{c}_1^{(i)} = 30\%$ in (b). As an illustration, we show the distributions of $c_1^{(i)}$ and $\tilde{c}_1^{(i)}$ over the whole population of taxpayers in blue and in red respectively. The distributions are not to scale.

varies dynamically depending on the taxpayer’s perceived deterrence level $\tilde{x}_t^{(i)}$ at time $t$. As the perceived deterrence level becomes very large, $\tilde{c}_1^{(i)}$ is assumed to approach $c_2$ asymptotically. As before, when taxpayer’s $i$ effective tax rate $T_t^{(i)}$ exceeds the threshold value $\tilde{c}_1^{(i)}$, taxes are perceived to be too high and s/he will begin to consider under-reporting to the IRS. The taxpayer’s desire to comply is again assumed to fall linearly to $c_2$ with increasing effective tax rate. Figure 7.1b illustrates how the value of $\tilde{c}_1^{(i)}$ may be distributed in the population. It also shows how the fairness function of a taxpayer changes and becomes a function describing his/her motivation to comply when his/her perceived deterrence is non-zero. For example, in our illustrated case shown in Figure 7.1, we have that with no perceived deterrence, a taxpayer’s $c_1^{(i)}$ value is equal to 15\%, and with some non-zero value of perceived deterrence his/her $\tilde{c}_1^{(i)}$ value is larger and equals 30\%. Thus, in presence of perceived deterrence $\tilde{x}_t^{(i)}$, a taxpayer’s compliance behavior is described by what we refer to as a motivation to comply function. This function is equal to the taxpayer’s fairness function when his/her perceived deterrence rate is zero. We assume that the motivation to comply function can be approximated as a piece-wise linear function of the effective tax rate $T_t^{(i)}$ and expressed as

$$F(T_t^{(i)}, \tilde{x}_t^{(i)}) = \begin{cases} 1 & \text{if } T_t^{(i)} < \tilde{c}_1(\tilde{x}_t^{(i)}); \\ \frac{[c_2 - T_t^{(i)}]}{[c_2 - \tilde{c}_1(\tilde{x}_t^{(i)})]} & \tilde{c}_1(\tilde{x}_t^{(i)}) \leq T_t^{(i)} \leq c_2; \\ 0 & \text{if } T_t^{(i)} > c_2. \end{cases}$$

(7.1)

The value of $\tilde{c}_1(\tilde{x}_t^{(i)})$ is calculated as

$$\tilde{c}_1(\tilde{x}_t^{(i)}) = c_1^{(i)} + (c_2 - c_1^{(i)})\Phi(\tilde{x}_t^{(i)}),$$

(7.2)

where $\Phi(\tilde{x}_t^{(i)})$ represent an S-Curved function that ranges between in $[0,1]$. Since $\Phi(\tilde{x}_t^{(i)})$ is an increasing function of the perceived deterrence $\tilde{x}_t^{(i)}$, as the perceived deterrence increases, the value of $\tilde{c}_1(\tilde{x}_t^{(i)})$ increases accordingly. In our model, we chose $\Phi(\tilde{x}_t^{(i)})$ to take the functional form.
Perceived Deterrence $\tilde{x}(i) = \tilde{q}(i) \tilde{P}(i)$

Figure 7.2: The form for the S-Curve $\Phi(x, m_x, s_x)$ which ranges in $[0, 1]$. As can be seen by the scale on the right hand side of the figure, the value for $\tilde{c}_1(\tilde{x}(i))$ increases from $c_1^{(i)}$ to $c_2$ as the perceived deterrence increases.

of a log-normal cumulative distribution function which can be expressed as

$$\Phi(x, m_x, s_x) = \frac{1}{2} + \frac{1}{2} \text{erf}\left[\frac{s_x(\ln x - \ln m_x)}{\sqrt{2}}\right], \quad (7.3)$$

where the parameter $m_x$ is the value of the perceived deterrence such that $\Phi(m_x) = 0.5$ and consequently it is the perceived deterrence point where $\tilde{c}_1(\tilde{x}(i))$ is half-way between $c_1^{(i)}$ and $c_2$. Figure 7.2 provides an illustration of this function. The parameter $s_x$ controls the steepness of the curve. Specifically, at $x = m_x$ the gradient of the curve is given by $g_x = s_x / (\sqrt{2}\pi m_x)$. The form of the function $\Phi(x, m_x, s_x)$ plays a key role in our taxpayer’s behavioral model because it controls how the taxpayer’s propensity to fully comply changes with perceived deterrence. We assume that the values of $m_x$ and $s_x$, and thus the form $\Phi(x, m_x, s_x)$ is the same for all taxpayers.

The second component that taxpayers factor into their personal evaluation is whether or not they received a tax refund from the IRS in the previous year. The personal evaluation of taxpayers is given by a weighted sum of $F(T_t^{(i)}, \tilde{x}_t^{(i)})$ and an indicator variable $R_t^{(i)}$ that specifies whether taxpayer $i$ received a tax refund in year $t$. Hence,

$$\Delta_P^{(i)} = r_w R_t^{(i)} + (1 - r_w) F(T_t^{(i)}, \tilde{x}_t^{(i)}), \quad (7.4)$$

where $r_w$ is a weight that provides the relative importance of this effect with respect to fairness and other motivations to comply considerations. Our ABM allows us to selected whether we want to include the effect from a tax refund. When this effect is unselected, $\Delta_P^{(i)}$ is simply set equal to

\footnote{Note that although two taxpayers may have the same perceived deterrence they will not necessarily have the same $c_1^{(i)}$. This is because they may still have different starting $c_1^{(i)}$ values. Thus, heterogeneity in their response to changes in perceived deterrence rate is due to their heterogeneity in their starting $c_1^{(i)}$ value. Those with a lower $c_1^{(i)}$ value are more responsive to an increase in perceived deterrence rate.}
\[ F(T^{(i)}_t, \hat{x}^{(i)}_t) \]

Noncompliant taxpayers that were not caught and penalized by the IRS have a strong motivation to continue to under-report their income to the IRS in the following year. The longer they have evaded their taxes at the same or greater level, the more likely they will continue to under-report to the IRS. In our taxpayer’s behavioral model (TBM), we initially modeled this effect by stochastically assigning the personal-level evaluation \( \Delta^{(i)}_p \) as follows

\[
\Delta^{(i)}_p = r_w R^{(i)}_t + (1 - r_w) \cdot \begin{cases} 
F(T^{(i)}_t, \hat{x}^{(i)}_t) & \text{with probability } [w^{(i)}_{t-1}]^v \\
0 & \text{with probability } 1 - [w^{(i)}_{t-1}]^v
\end{cases}
\] (7.5)

This stochastic assignment makes sure that under-reporting taxpayers that have gone through many years without being audited and penalized will tend towards full evasion at a faster rate than those who have only just begun to under-report. The exponent parameter \( v \) controls how fast those that initiate in under-reporting tend towards full evasion. The larger the exponent \( v \) the sooner a taxpayer who initiated in under-reporting will become a full tax evader. This stochastic assignment does allow for the possibility for a taxpayer who recently initiated in under-reporting to return to be fully compliant even without being audited. However, this is more likely to happen when the taxpayer’s evaluation of \( F(T^{(i)}_t, \hat{x}^{(i)}_t) \) is close to 1 and the number of years since s/he last initiated in under-reporting is small. Appendix \( A \) provides an analysis of the effect of this stochastic assignment on the EWMA process for the dynamics of \( w^{(i)}_t \).

However, we found that the stochastic assignment described in equation 7.5 proved to be more complicated than needed a similar effect and related results are produced by the following assignment of \( \Delta^{(i)}_p \) to those taxpayers that have already initiated in tax evasion. This simplified rule assigned the value of \( \Delta^{(i)}_p \) to 0 for all evading taxpayers. However, if the evading taxpayer is not a long-time evader and has only evaded on his/her taxes for a duration characterized by a timescale that is equal to the taxing generation half-life \( \tau_G \), then s/he has may spontaneously regret evading and has a possibility of becoming compliant again. This happens as long as his/her value for \( F(T^{(i)}_t, 0) \) describing the taxpayer’s fairness function (i.e., whether s/he perceives the effective tax rate to be fair) is equal to 1. Therefore, similarly to the process described by equation 7.5 an evading taxpayer can spontaneously become compliant again without IRS intervention as long as they haven’t become long-term tax evaders and perceive the effective tax rate that they are asked to pay to be fair. Although we implemented both versions of the model with the two different assignments, we chose to use this simplified version. This is because the simplified version produces the same qualitative effects and removes the need for the additional parameter \( v \) used in equation 7.5.

7.2 The calculation of the effect of the penalties \( \Delta^{(i)}_{\text{Penalty}} \)

Taxpayers that are caught for tax evasion and penalized by the IRS calculate a quantity \( \Delta^{(i)}_{\text{Penalty}} \) as follows

\[
\Delta^{(i)}_{\text{Penalty}} = \mathcal{E}^{(i)}[\mathcal{E}_A] \sum_{m=0}^{K} \frac{[T^{(i)}_{t-m}(I^{(i)} - A^{(i)}_{t-m})(1 + P_{t-m})]}{T^{(i)}_{t-m} I^{(i)}} = \mathcal{E}^{(i)}[\mathcal{E}_A] \sum_{m=0}^{K} [u^{(i)}_{t-m}(1 + P_{t-m})]
\] (7.6)

where \( u^{(i)}_t = (I^{(i)} - A^{(i)}_t)/I^{(i)} \) is the income amount that taxpayer \( i \) under-reported in year \( t \) expressed as a proportion of his/her income and \( \mathcal{E}^{(i)}[\mathcal{E}_A] \) represents a stochastic realization of how
efficient the tax audit was in detecting the unreported income of taxpayer $i$. Here, $\epsilon_A$ denotes the mean detection efficiency and $\mathcal{E}$ denotes the sampling distribution function. We use a uniform distribution for $\mathcal{E}$, such that the detection efficacy is bounded between $\pm 10\%$ of the mean detection efficiency $\epsilon_A$. The denominator in equation (7.6) represents the full amount of yearly taxes that are due to the IRS by taxpayer $i$. The numerator in equation (7.6) is equal to the total past taxes due to the IRS plus the penalties by the evading taxpayer $i$. When summed over all taxpayers, this quantity gives the amount of recovered taxes and penalties. We define this as

$$R_{\text{Penalty}} = \sum_{i \in I^p_t} \mathcal{E}(i)[\epsilon_A] \sum_{m=0}^K [T_{i-m}^m(I^i - A_{i-m}^i)(1 + P_{i-m})],$$

where $I^p_t$ represents the set of taxpayers that were caught and penalized for tax evasion in year $t$. Once a taxpayer is caught and penalized for tax evasion for a past tax year $m$, his/her value for $A_{i-m}^i$ is set equal to $T_{i-m}^m$. This prevents taxpayers that are audited in two subsequent consecutive years to be penalized twice for having evaded in tax year $m$.

In Section 6.2 we mentioned that the value for $\Delta_p(i)$ of taxpayers that are caught for tax evasion and penalized by the IRS is $\phi \Delta_{\text{Penalty}}^i$. The scaling parameter $\phi$ controls the impact of the evaluation of the penalty $\Delta_p(i)$ on the taxpayer’s propensity to relapse back into tax evasion behavior, i.e., $w_{t+1}$. Consider a taxpayer that is caught for tax evasion and penalized by the IRS and his/her value for $\Delta_{\text{Penalty}}^i$ is $100\%$. We can specify that after being penalized this taxpayer will have a propensity $w_{t+1}$ to stay compliant and not relapse into tax evasion behavior that is at least equal to a specified value $\vartheta$. So by enforcing that $w_{t+1} \geq \vartheta$ and by following the mathematical analysis of EWMA presented in appendix A, Section A.2 we find that $\phi = sN_{t+1}/(1 - \vartheta)$ where $N_{t+1}$ is the normalizing factor representing the maximum value $V_{t+1}$ could be in year $t+1$.

A natural choice for the tuning parameter value $\vartheta$ is $\vartheta = (s + 1)^{-1}$ which leads to $\phi = N_{t+1}$. In this case and for $\Delta_{\text{Penalty}}^i = 100\%$, a penalized taxpayer’s future compliance behavior will be equally affected by the penalty as his/her whole history of past evaluations up to year $t$. If instead $\Delta_{\text{Penalty}}^i = 200\%$, his/her compliance behavior will be twice more affected by the penalty as his/her whole history of past evaluations up to year $t$.

### 7.3 The calculation of the media feedback term on tax morale and fairness $\Delta_M$

The media relevant taxpayer’s evaluation $\Delta_M$ depends on what the media reports about the problem of tax evasion and the level of the tax gap. We assume that $\Delta_M$ depends on the gross tax gap $X_{t}^{\text{[gross]}}$ as follows

$$\Delta_M(X_{t}^{\text{[gross]}}) = 1 - \Phi_M(X_{t}^{\text{[gross]}}),$$

where

$$\Phi_M(X_{t}^{\text{[gross]}}), m_m, s_m) = \frac{1}{2} + \frac{1}{2} \text{erf} \left[ \frac{s_m(\ln X_{t}^{\text{[gross]}} - \ln m_m)}{\sqrt{2}} \right],$$

depends on the gross tax gap according to a log normal S-Curve, and where $m_m$ and $s_m$ are fitting parameters. In the hypothetical case that $X_{t}^{\text{[gross]}}$ is equal to $100\%$, the value of $\Delta_M$ is zero and hence

---

11We emphasize that $\mathcal{E}(i)[\epsilon_A]$ is not the efficacy of determining taxpayer’s $i$ income but rather the efficacy of detecting his/her income that is unreported to the IRS.

12It may appear that the model is very sensitive to the value of the parameter $\phi$ and hence $\vartheta$. However, we found that as long as $\vartheta$ is about $\sim (s + 1)^{-1}$ or larger, the model is not so sensitive to this parameter.
the taxpayers are less willing to be compliant. If instead the gross tax gap is $\Xi^{(\text{gross})} = m_m$, then $\Delta_M = 0.5$.

### 7.4 The effect of the media feedback on the taxpayer’s perceived audit rate $\tilde{q}_t^{(i)}$

Media tracks a moving average value of two indicators over a time window that includes the past $\tau^{(\text{Media})}$ years. The first is the average yearly number of audits denoted by $\bar{q}_t^{(M)}$. The second is the average yearly total dollar amount of recovered taxes and penalties denoted by $\bar{P}_t^{(M)}$. If in the current year either the total number of audits or total dollar amount of recovered taxes and penalties is significantly greater than their moving average, the media reports to the taxpayers the news of the extraordinary activities levels by the IRS in auditing and/or catching tax evaders, potentially modifying taxpayers perceived risks of being audited by the IRS in future years. We use a simple statistical Z-test and a 99% confidence threshold level to determine whether the difference is significant.

However, in our model, most taxpayers will only pay attention to this news stochastically with a probability $\zeta_t^{(i)}$. To determine whether the taxpayer pays attention to the news we first calculate the income deciles of all the taxpayers that were selected for an audit by the IRS. Then, for each taxpayer, we find in which of these income deciles they fall into. For any given taxpayer, the particular decile s/he falls into provides the percentage of taxpayers that were audited by the IRS who have similar or lower incomes. This percentage is interpreted in our model as the probability $\zeta_t^{(i)}$ that the taxpayer pays attention to media reports of extraordinary activities levels by the IRS in auditing and/or catching tax evaders. However, some taxpayers will always pay attention when media report this news, irrespective of the breakdown in income of who was audited. These taxpayers are identified as those who have an active (i.e., true) value for their actor attribute and their value for the probability $\zeta_t^{(i)}$ is always 1.

### 7.5 The calculation for the taxpayer’s perceived audit rate $\tilde{q}_t^{(i)}$

A taxpayer’s perceived deterrence $\tilde{x}_t^{(i)}$ is given by the product of his/her perception of the risk of being audited $\tilde{q}_t^{(i)}$ and his/her perception of the penalty rate $\tilde{P}_t^{(i)}$. In our model, each taxpayer has an initial baseline perception of the risk of being audited by the IRS. For taxpayer $i$ this initial risk perception is denoted by $\tilde{q}_0^{(i)}$. Over time this perception, denoted by $\tilde{q}_t^{(i)}$ can change due to (i) personal experiences of being audited, (ii) communications/observations of alters being audited by the IRS and (iii) media feedback which report whether the IRS has been carried out an unusually high number of audits and has caught many evaders. At the end of each year, our model evaluated three indicator variables for each taxpayer.

1. The first, $\mathcal{A}_p^{(i)}$ is equal to 1 if taxpayer $i$ was audited by the IRS at the end of the current year, and equal to 0 otherwise.

2. The second, $\mathcal{A}_N^{(i)}$ is equal to 1 if any of the alters of taxpayer $i$ were audited by the IRS at the end of the current year, and equal to 0 otherwise.

3. The third, $\mathcal{A}_M^{(i)}$ is equal to 1 if at the end of the current year media reports about the activity of the IRS and this news is of interest to taxpayer $i$, and equal to 0 otherwise. Section 7.4 described the conditions needed for $\mathcal{A}_M^{(i)} = 1$.  

35
The weighted sum, $\mathcal{A}_i^{(i)}$ given by

$$\mathcal{A}_i^{(i)} = \frac{\beta_P A_P^{(i)} + \beta_N A_N^{(i)} + \beta_M A_M^{(i)}}{\beta_P + \beta_N A_N^{(i)} + \beta_M A_M^{(i)}},$$

(7.10)

is used to updated taxpayer’s $i$ risk perception of being audited in the next year. Equation (7.10) takes the same form as equation (6.2). The denominator represents the normalization term and reflects the fact that taxpayers will always take into consideration whether they themselves were audited or not, while network and media feedback effects are only taken into consideration if they are exposed to them as determined by $A_N^{(i)}$ and $A_M^{(i)}$ respectively.

We assume that each taxpayer updates his/her total tallied number of audit experiences and observations using an EWMA as follows

$$T_{i+1} = s T_i + (1 - s) A_i^{(i)}.$$  

(7.11)

In absence of network or media effects, $T_i$ can be interpreted as the number of times that taxpayer $i$ experienced an audit using an EWMA to discount the importance of past audit events. The perceived risk of being audited $\tilde{q}_i^{(i)}$ linearly depends on $T_i$ and is expressed as

$$\tilde{q}_i^{(i)} = \tilde{q}_0^{(i)} + (1 - \tilde{q}_0^{(i)}) T_i.$$  

(7.12)

Thus, when $T_i = 0$ the perceived risk of being audited $\tilde{q}_0^{(i)}$ is equal to the baseline value $\tilde{q}_0^{(i)}$, whereas in a hypothetical and extremely unlikely case that $T_i = 1$ which can occur if the taxpayer is audited every year, his/her perceived risk of being audited $\tilde{q}_i^{(i)}$ is equal to 100%. Therefore, equation (7.12) bounds the perceived audit rate in the range $[\tilde{q}_0^{(i)}, 1]$ irrespective of the true audit rate.

This formulation does not allow for the perceived audit rate to decrease below the baseline $\tilde{q}_0^{(i)}$ even if the true audit rate was drastically reduced. To correct for this effect, we assume that the distribution for the baseline perceived audit rate in the population $\tilde{q}_0^{(i)}$ linearly decreases if the true audit rate is decreased from its initial status-quo value. Specifically, when the average audit rate $q_i$ used by the IRS is decreased from its initial status-quo value by a certain proportion $\xi$, we allow the baseline perceived audit rates $\tilde{q}_0^{(i)}$ for each taxpayer to exponentially decay towards $\xi \tilde{q}_0^{(i)}$ as follows

$$\tilde{q}_i^{(i)} = s (\tilde{q}_i^{(i)} - \xi \tilde{q}_0^{(i)}) + \xi \tilde{q}_0^{(i)}.$$  

(7.13)

### 7.5.1 The Bomb Crater and Gambler’s Fallacy Effects

Our model considers two additional effects that can modify the perceived risk of being audited for some taxpayers. The first of these is called the Bomb Crater Effect and is similar to the popular conception that “lightning never strikes twice in the same place”. Hence, taxpayers who have recently been audited may perceive that the IRS will not audit them again for the foreseeable near future especially if they were not caught evading on their taxes. We assume that $\tilde{q}_i^{(i)}$ is reduced by a multiplicative factor that depends on the number of years $a_i^{(i)}$ since taxpayer $i$ was last audited
by the IRS. The way \( \tilde{q}_{t+1}^{(i)} \) is modified due to this effect can be expressed as

\[
\tilde{q}_{t+1}^{(i)} \rightarrow \tilde{q}_{t+1}^{(i)} \{ 1 - \kappa_{\text{BC}} \exp[-(1-s)\alpha_{t}^{(i)}] \},
\]

where the multiplicative coefficient \( \kappa_{\text{BC}} \) is a positive constant. Thus, over time the multiplicative factor that modifies the perceived risk of being audited decays exponentially back to 1 from below.

The second of these is the Gambler’s Fallacy Effect and it temporarily increases the perceived risk of being audited by taxpayers that in recent years have started to under-report on their taxes. Taxpayers may believe that the risk of being selected for an audit by the IRS increases with (i) the number of years that pass without being audited and (ii) the larger the amount they under-report on their income. If the selection of who gets audited by the IRS were truly a random and memoryless process then this is a misconception of chance (i.e., a fallacy). We assume that \( \tilde{q}_{t+1}^{(i)} \) is increased by a multiplicative factor that depends on the number of years \( y_{t}^{(i)} \) since taxpayer \( i \) began to underreport. The way \( \tilde{q}_{t+1}^{(i)} \) is modified due to this effect can be expressed as

\[
\tilde{q}_{t+1}^{(i)} \rightarrow \tilde{q}_{t+1}^{(i)} \{ 1 + \kappa_{\text{GF}} \exp[-(1-s)y_{t}^{(i)}] \},
\]

where the multiplicative coefficient \( \kappa_{\text{GF}} \) is a positive value that depends on the proportion of the income that taxpayer \( i \) reports to the IRS. Thus, over time the multiplicative factor that modifies the perceived risk of being audited decays exponentially back to 1 from above. This can be expressed as

\[
\kappa_{\text{GF}}^{(i)} = \exp[\gamma_{\text{GF}} \log(1 - \alpha_{t}^{(i)}) + \alpha_{\text{GF}}] - 1,
\]

where the constants \( \alpha_{\text{GF}} \) and \( \gamma_{\text{GF}} \). The larger the amount underreported income to the IRS the larger the coefficient \( \kappa_{\text{GF}}^{(i)} \). In our model, this effect slows down the rate at which a taxpayer who has recently initiated in tax evasion approaches the behavior of fully evading on his/her hideable income.

7.5.2 The dependence of the perceived audit rate on the true audit rate

The dynamics of how taxpayers change their perceived audit rate is described by an EWMA process formulated by Equations 7.10 to 7.13. The dynamics strongly depends on the starting baseline perceived audit rate \( \tilde{q}_{0}^{(i)} \). The blue curve in Figure 7.3 shows how the average perceived audit rate of the taxpayers in our ABM changes with the audit rate. Here, we consider just the EWMA process and exclude the bomb crater, gambler’s fallacy, and all media effects. Thus here, taxpayers update their perceived audit rate based on personal experiences and network effects only. The blue area or bands show the interquartile ranges of the perceived audit rate. The distribution of our taxpayer’s baseline perceived audit rates \( \tilde{q}_{0}^{(i)} \) was informed by the American Life Panel (ALP) survey that we fielded in 2016 which is discussed in Section 13. At baseline, the average effective audit rate over all taxpayers that we considered was 0.8% which is representative of the audit rate in 2016 when the ALP survey was fielded. Figure 7.3 shows that the mean perceived audit rate has a nearly linear dependence on the audit rate. This can be explained by the linear dependence of the expectation value of the personal experiences indicator variable \( A_{t}^{(i)} \) with the audit rate. The deviation from the linear dependence is due to the network effects and the dependence of the
Figure 7.3: Plot showing the dependence of the mean perceived audit rate changes on the actual audit rate using a linear scale (a) and log-log scale (b). The curve shown in blue shows how the mean perceived audit rate of the taxpayers in our ABM changes with the actual audit rate. The blue area or bands show the interquartile ranges of the perceived audit rate. The vertical dotted line on the base 10 log-log scale plot shows an audit rate of 0.8% which represents the audit rate in 2016 when the ALP survey was fielded. The interquartile ranges of the perceived audit rate at the vertical dotted line correspond to those found from our ALP survey. The curve shown in red provides a smoothed fit to the functional dependence averaged over previous studies that are referred to in the main text. The red area or band provides the range reported in the different studies. The green curve and bands shown in the log-log scale plot show how these curves considered in the literature extend to very small audit rates by assuming a third order polynomial fit.

Previous studies have considered and modeled this functional dependency \[34, 169, 183\]. These studies have been summarized by Bloomquist (2011) \[23\]. Figure 7.3 shows how this dependence that is produced by our model dynamics compares to the predictions made by these previous studies. We note that in the baseline case when the audit rate is 0.8% the difference between what is produced by our model and what is predicted by previous studies is very large. However, this is entirely due to our starting distributions for the baseline perceived audit rate as found by the ALP survey. As reported in Section 13, respondents in our survey had very large perceived audit rates.

### 7.6 The calculation for the taxpayer’s perceived penalty rate \( \tilde{P}_t^{(i)} \)

Each taxpayer has an initial baseline perception of the penalty rate applied by the IRS on unpaid taxes. For taxpayer \( i \) this initial perceived penalty rate is denote by \( \tilde{P}_0^{(i)} \). Since the official penalty rate is more easily observed than audit rates, over time the perceived penalty rate is updated in a different way than the perceived audit rate. Being penalized for unpaid taxes is a salient event where taxpayers learn about the true penalty rate and consequently are unlikely to forget it for the rest of their taxpaying lifetime. Even if a taxpayer is simply audited and is not caught by the IRS for unpaid taxes, s/he could become aware of the true penalty rate. In addition to being audited by the IRS, there is a chance that taxpayers can learn about the true penalty rate from their alters being audited or from the media. Therefore, we assume that a taxpayer learns about the true penalty rate with probability \( A_t^{(i)} \). Notice that if the taxpayer is audited by the IRS

\[ \text{expected probability of at least one alter being audited} = 1 - (1 - q)^k, \]

\( k \) here represents the number of alters of taxpayer \( i \).
and possibly penalized for tax evasion, this probability is 100%. Therefore, with probability $A^{(i)}_t$ a taxpayer perceived penalty rate changes abruptly from its current value $\overline{P}^{(i)}_t$ to the actual value for the penalty rate used by the IRS (i.e., $P$). In subsequent years, the value of the perceived penalty slowly decays back to $\overline{P}^{(i)}_0$ according to an EWMA. In our model, this decay occurs over a very long period given by a typical taxpayer generation time. Thus, the expression describing how the taxpayer $i$ updates his/her perceived penalty rate is

$$
\overline{P}^{(i)}_{t+1} = \begin{cases} 
P & \text{with probability } A^{(i)}_t \\
S_G \overline{P}^{(i)}_t + (1 - S_G) \overline{P}^{(i)}_0 & \text{with probability } 1 - A^{(i)}_t
\end{cases}
$$

(7.17)

where $S_G = e^{-\log(2)/\tau_G}$ and $\tau_G$ is the taxpaying generation half-life.

8 Specifications and Initialization of the IRS Model Components

8.1 The effective tax rate

Table 8.1 show the US tax brackets for 2016.\footnote{Data for 2016 tax brackets is available at \url{https://taxfoundation.org/2016-tax-brackets/}. Historical tax brackets are available at \url{http://taxfoundation.org/article/us-federal-individual-income-tax-rates-history-1913-2013-nominal-and-inflation-adjusted-brackets}. The model can also read and use the proposed changes to the tax brackets and marginal tax rates proposed by President Trump which can be found at \url{http://www.taxpolicycenter.org/sites/default/files/publication/135696/2000983-Families-Facing-Tax-Increases-Under-Trumps-Plan.pdf}} We used the filing status of each taxpayer and their income to calculate the effective tax rate $T^{(i)}$ for each taxpayer $i$. In our model, the governing state (GS) can change its fiscal policy, the tax brackets and the marginal tax rates that are applied to each tax bracket. However, for the purpose of model validation and calibration the fiscal policy is assumed to not change over time and thus each taxpayer is taxed at the same effective tax rate each year.

<table>
<thead>
<tr>
<th>Marginal Tax Rate (%)</th>
<th>Head of Household ($k$)</th>
<th>Married Filing Jointly ($k$)</th>
<th>Married Filing Separately ($k$)</th>
<th>Single ($k$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>[0, 13.25)</td>
<td>[0, 18.85)</td>
<td>[0, 9.275)</td>
<td>[0, 9.275)</td>
</tr>
<tr>
<td>15</td>
<td>[13.25, 50.4)</td>
<td>[18.85, 75.3)</td>
<td>[9.275, 37.65)</td>
<td>[9.275, 37.65)</td>
</tr>
<tr>
<td>25</td>
<td>[50.4, 130.15)</td>
<td>[75.3, 151.9)</td>
<td>[37.65, 75.95)</td>
<td>[37.65, 91.15)</td>
</tr>
<tr>
<td>28</td>
<td>[130.15, 210.8)</td>
<td>[151.9, 231.45)</td>
<td>[75.95, 115.725)</td>
<td>[91.15, 190.15)</td>
</tr>
<tr>
<td>33</td>
<td>[210.8, 413.35)</td>
<td>[231.45, 413.35)</td>
<td>[115.725, 206.675)</td>
<td>[190.15, 413.35)</td>
</tr>
<tr>
<td>35</td>
<td>[413.35, 441)</td>
<td>[413.35, 466.95)</td>
<td>[206.675, 233.475)</td>
<td>[413.35, 415.05)</td>
</tr>
<tr>
<td>39.6</td>
<td>[441, 1e+06)</td>
<td>[466.95, 1e+06)</td>
<td>[233.475, 1e+06)</td>
<td>[415.05, 1e+06)</td>
</tr>
</tbody>
</table>

Table 8.1: US Tax brackets for 2016.
8.2 The transitions between a tax refund and balance due

According to the IRS data, nearly eight out of ten taxpayers get a federal tax refund every year. The actual proportion varies by the taxpayer’s income and employment type [32]. Using this data, an analysis by H&R Block’s Tax Institute, and assuming a simple Markov process, we calculated the transition rates \( \rho_{a \rightarrow r} \) and \( \rho_{r \rightarrow a} \) by income reported bracket [15]. Assuming an equilibrium state, for a given income bracket the ratio \( \rho_{r \rightarrow a} / \rho_{a \rightarrow r} \) is equal to \( (1-r)/r \) where here \( r \) represents the proportion for taxpayers that receive a tax refund. Thus, \( \rho_{r \rightarrow a} = \rho (1-r) \) and \( \rho_{a \rightarrow r} = \rho r \), where \( \rho \) is an overall transition rate and a parameter of our model.

8.3 The initial IRS audit strategy and rate

As our initial audit strategy, the IRS assigns a probability of being selected for an audit to each taxpayer which is calculated according to the income a taxpayer reports to the IRS and on his/her suspected proportion of hideable income. To inform these probabilities, we used IRS data from 2015 that is summarized in Table 8.2 that provides the operational audit rate applied to different AGI brackets [16]. Since in our model the proportion of taxpayers belonging to each income bracket does not change over time, we can assume that the audit rates applied to each AGI bracket shown in Table 8.2 scales linearly with respect to the average effective audit rate. Thus, we can find how the audit rate for each AGI bracket changes as we change the overall audit rate at the population-level. We further assume that within each AGI bracket the IRS preferentially selects the self-employed or those with a higher suspected proportion of hideable income for an audit. Since by definition all the income of a self-employed is hideable, we chose to use the proportion of hideable income attribute as our selection weights within each bracket.

To inform the initial IRS audit rate and its variability we used IRS data for years 2006 to 2015 [17]. The audit rate in 2015 was 0.84%. However, the average audit rate over these years was 0.99% and the coefficient of variability was 9.25%. We used these numbers to stochastically sample the total number of audits from a normal distribution. Our model assumes that the 9.25% coefficient of variability is a fixed constant and we calculate the variance of the normal distribution of our stochastic sampling based on the audit rate.

8.4 IRS Fixed Operational Costs and Variable Audit Costs

Based on Table 28 of the IRS Data Book for the fiscal year 2016 [18], the enforcement budget for collections and enforcement was $3.948 billion. This amount includes spending for collection, appeals, and counsel. Table 30 of the IRS Data Book shows there were 33,426 full-time equivalent (FTE) positions in enforcement and collection. The same table shows 10,174 FTE positions for revenue agents, 8,267 FTE positions for tax examiners, and 1,227 FTE positions for tax technicians. This provides an approximate estimate of FTE positions for purely audit activities and excludes revenue officers (collection), appeals, and attorneys. The sum of these three FTE positions for

---


Table 8.2: Audit rates for different income brackets.

<table>
<thead>
<tr>
<th>Income ($)</th>
<th>Audit Rate (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non filers</td>
<td>3.78</td>
</tr>
<tr>
<td>(0, 25)</td>
<td>1.01</td>
</tr>
<tr>
<td>[25, 50]</td>
<td>0.50</td>
</tr>
<tr>
<td>[50, 75]</td>
<td>0.47</td>
</tr>
<tr>
<td>[75, 100]</td>
<td>0.49</td>
</tr>
<tr>
<td>[100, 200]</td>
<td>0.64</td>
</tr>
<tr>
<td>[200, 500]</td>
<td>1.54</td>
</tr>
<tr>
<td>[500, 1,000]</td>
<td>3.81</td>
</tr>
<tr>
<td>[1,000, 5,000]</td>
<td>8.42</td>
</tr>
<tr>
<td>[5,000, 10,000]</td>
<td>19.44</td>
</tr>
<tr>
<td>10,000+</td>
<td>34.69</td>
</tr>
</tbody>
</table>

purely audit activities as a proportion of the total FTE positions in IRS enforcement and collection is 59%. Thus, we can assume that $2.329 billion (i.e., 59% of $3.948 billion) represent just the audit-related expenditures of enforcement and collection. Since in the fiscal year 2016 there were 1.035 million audits, we calculated an estimated cost per audit of $2,250. However, for our model, our interest is in how the cost depends on income and on the type of audit. Table 9a of the IRS Data Book presents the number of field and correspondence audits for different income brackets. We used this table to estimate the audit cost by income and by type of audit. Table 8.3 provides our estimates for the average cost per audit.

Table 8.3: Table showing our estimates for the average cost per audit broken by income and by type of audit. The table shows the estimated average cost of an audit for field and correspondence audits. The last column shows the weighted average estimated cost per audit.

<table>
<thead>
<tr>
<th>Income Range ($ k)</th>
<th>Field Audit Cost ($)</th>
<th>Correspondence Audit Cost ($)</th>
<th>Audit Cost ($)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(0, 25)</td>
<td>1,977</td>
<td>1,293</td>
<td>1,744</td>
</tr>
<tr>
<td>[25, 100)</td>
<td>2,506</td>
<td>2,230</td>
<td>2,395</td>
</tr>
<tr>
<td>[100, 200)</td>
<td>5,781</td>
<td>3,507</td>
<td>5,166</td>
</tr>
<tr>
<td>[200, 1,000)</td>
<td>7,627</td>
<td>3,015</td>
<td>10,642</td>
</tr>
<tr>
<td>1,000+</td>
<td>28,453</td>
<td>13,078</td>
<td>41,531</td>
</tr>
</tbody>
</table>

As a model validation step, we found that the model produces an average cost per audit of $2,450. This output is very similar to the value of $2,250 we estimated from the IRS tables. The model validation considered the audit costs given in Table 8.3 and applied the audit strategy described in Section 8.3. We further assumed that audited taxpayers with the highest incomes would be selected for a field audit and that field audits account for 25% of all audits. This validation provides some indication that the way taxpayers are selected for an audit by the model produces the same results as the IRS.
same average cost per audit as the one calculated by IRS tables.

9 Specifications of the Governing State (GS) Components

Our conceptualized model includes the Governing State (GS). The GS can change its fiscal policy by changing the tax brackets and the applied marginal tax rates in its attempt to minimize tax evasion, yearly deficits, and national debt while providing for public services to the taxpayers. It can also increase funding to the IRS to intensify and enhance their deterrence activity. However, our current model implementation described in Sections 6 and 7 does not describe the GS. This is because our first focus of the model implementation was to validate and calibrate the model under the status-quo conditions that consider the current fiscal and deterrence policies without varying them over time. In this section, we describe the module for the GS and how it uses the aggregated compliance outputs from the model to update national indicators for the deficit and the debt. Throughout the section, we refer to Table 9.1 which shows a list of US national level macroeconomic quantities which our model either uses as inputs or as targets. It also shows model outputs from our calibrated best case that have been scaled up to be representative of at the national level. The calibration of the model is discussed in Section 20. We use the table to describe how close our model can reproduce various macro-level targets.

<table>
<thead>
<tr>
<th>input</th>
<th>target</th>
<th>model</th>
<th>scaled.model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Revenues (trillion $)</td>
<td>3.277</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Income Tax Revenues (trillion $)</td>
<td>1.447</td>
<td>1.664</td>
<td>1.447</td>
</tr>
<tr>
<td>Income Tax Revenue per tax payer (thousand $)</td>
<td>9.583</td>
<td>11.020</td>
<td>9.583</td>
</tr>
<tr>
<td>Percentage of Other Revenues (%)</td>
<td>56.00</td>
<td>48.51</td>
<td>55.31</td>
</tr>
<tr>
<td>Tax Gap (%)</td>
<td>16.30</td>
<td>17.01</td>
<td>17.01</td>
</tr>
<tr>
<td>Recovered Tax Revenues (billion $)</td>
<td>28.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Recovered Tax Revenues and Penalties (billion $)</td>
<td>45.348</td>
<td>39.433</td>
<td></td>
</tr>
<tr>
<td>Gov. Expenses (trillion $)</td>
<td>3.863</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fixed Gov. Expenses (trillion $)</td>
<td>3.857</td>
<td>3.857</td>
<td></td>
</tr>
<tr>
<td>Audit Costs (billion $)</td>
<td>2.329</td>
<td>2.960</td>
<td>2.960</td>
</tr>
<tr>
<td>Mean Cost per Audit (thousand $)</td>
<td>2.250</td>
<td>2.450</td>
<td>2.450</td>
</tr>
<tr>
<td>Deficit (billions $)</td>
<td>-586</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 9.1: Table showing macro-level economic quantities at the US national-level that are used as either inputs or targets as shown by the first two columns for our set of calibrated model runs. For our single calibrated best run case, the third column shows our model outputs when scaled and interpreted at the national level. The fourth column illustrates how we scale our model outputs to correct for the mismatch in total individual inline tax revenues shown in the second row.

In 2016 the US population was 324 million and the total number of individual income tax returns were 151 million. As shown in Table 9.1 the US total revenues were $3.277 trillion. The total tax revenue per capita was $10,114, total per capita and spending was $11,923 leading to a deficit of $1,809 per capita. Therefore, the total revenues per individual income tax return

---

20 From IRS data book Table 2 for 2016 can be found at https://www.irs.gov/pub/irs-soi/16databk.pdf
21 Data obtained at http://www.usgovernmentspending.com/year_spending_2016USdn_18da2n_3OF0G0G2#.usgs302 The per capita spending and revenue can also be estimated with the data found at https://www.fiscal.treasury.gov/fsreports/rpt/combStmt/cs2016/receipt.pdf and https://www.fiscal.treasury.gov/fsreports/rpt/combStmt/cs2016/outlay.pdf Also, the U.S. Census Bureau, Population Division, Table 1.
(i.e., per taxpayer) was equal to $R = $21,702. Revenues to the GS come from the total income taxes and penalties paid by the taxpayers as well as other government revenues such as payroll taxes, corporate income taxes, excise taxes, the estate tax, and other taxes and fees. In 2016 the IRS collected $1.447 trillion in net income tax and $9,580 per taxpayer. Based on these numbers, this represented 44% of total revenues per taxpayer. Therefore, taken together the other government revenues represented 56% of the total government revenues in 2016.

9.1 The total revenues to the GS

In our model total revenues is equal to

$$Q_t = \sum_i T_t^{(i)} A_t^{(i)} + R_{\text{Penalty}} + Z_{\text{SQ}}$$  \hspace{1cm} (9.1)$$

where the first term represents the sum of the tax revenues collected from the taxes paid by the taxpayers in year $t$, the second term $R_{\text{Penalty}}$ is given by equation 7.7 and represents the recovered revenues and penalties by the IRS from the audit activities, and $Z_{\text{SQ}}$ represents other government revenues at the status quo. Neglecting recovered revenues from past taxes and penalties, in our model the constant amount $Z_{\text{SQ}}$ is set equal to

$$Z_{\text{SQ}} = \alpha_R N - \sum_i T_0^{(i)} A_0^{(i)}$$  \hspace{1cm} (9.2)$$

where $N$ represents the number of taxpayers in our model and the last term represents the initial, status-quo total income tax revenues produced by the model. For the purpose of our model, other government revenues $Z_{\text{SQ}}$ is calculated using equation 9.2 and using the 2016 value for $\alpha_R$ and is then fixed and considered a constant for the rest of the dynamics produced by the simulation.

Table 9.1 shows that our base case run produces total income tax revenue of $1.664 trillion which translates to $11,020 per taxpayer. Moreover, this represents 51.49% of the total US revenues which is listed as an input, and hence the other government revenues $Z_{\text{SQ}}$ expressed as a percentage of total revenues is 48.51%.

9.2 The Tax Gap

Here we refer to a 2016 IRS report on the annual averages of tax compliances. Between tax years 2008 to 2010, on average the true tax liability was estimated to be $2,496 billion. The average yearly taxes paid voluntarily and timely was $2,038 leading to a gross tax gap of $458 billion or 18.3% of the true tax liability. On average, $52 billion per year were recovered through IRS audit activities leading to a net tax gap of $406 billion or 16.3%. The reported gross tax gap includes many components. In addition to the income tax component, the reported tax gap includes corporate, employment, estate and excise taxes. For each of these components, the reported tax gap

---

22 Data found in Table 1 of https://www.irs.gov/pub/irs-soi/16databk.pdf The gross collections were $1.267 trillion in individual income tax withheld and $518 billion in individual income tax payments. However, the net collections were $1.447 trillion.

23 This number is slightly smaller than the 47% that is often quoted. For example, see https://www.fiscal.treasury.gov/fsreports/rpt/combsmt/c2016/receipt.pdf

24 See the 2016 IRS report on Tax Gap found at https://www.irs.gov/uac/the-tax-gap
includes non-filing, underreporting, and underpayment. For each component, the report lists the gross tax gap in billions of dollars and the amount recovered through IRS audit activities. For example, the gross tax gap that is specific to individual income taxes was $319 billion of which $28 billion is recovered through IRS audit activities. For corporate taxes, these numbers are $44 billion and $9 billion respectively. However, the report did not list the estimated true tax liability of each tax component and thus for individual income taxes, we do not know the gross tax gap expressed as a percentage of the true tax liability. Moreover, we do not know the tax gap that is specific to only those who under-report. Looking at the gross tax gap and the amount recovered through IRS audit activities for each component we can argue that compliance rate of individual income taxes component is the highest. Thus, although the gross tax gap expressed as a percentage of the true tax liability is 18.3% for all taxes, it may well be smaller when considering just the individual income taxes component.\footnote{This conclusion is in direct contrast to the Forbes article by Dr. Mathur that can be found at https://www.forbes.com/sites/aparnamathur/2016/05/27/a-simpler-tax-code-may-help-close-the-tax-gap. In this article, the average gross tax gap for individual income taxes was estimated to be 23% and that for corporate taxes 14%}

In our model, the average true tax liability per taxpayer is $13,278 which when scaled to the national level, corresponds to a true tax liability of $2,005 billion for the individual income taxes component. To reproduce a gross tax gap of $319 billion for individual income taxes we would thus require a gross tax gap of 15.9% of the true tax liability. Thus, as a percentage of true tax liability, our estimated gross tax gap for individual income taxes range between 15.9% and 18.3%. The average value between these two points is 17.1% which we use as our target gross tax gap. As shown in Table 9.1 and further discussed in Sections \[18\] and \[20\] our calibrated model reproduces this target value.

### 9.3 Scaling of the income tax revenues

Table 9.1 shows that our model best case run produces an individual income tax revenues of $1.664 trillion which although is within an acceptable range from our target value of $1.447 trillion, it is noticeably larger. This mismatch leads to a mismatch with our targets for two additional outputs, these are (i) the individual income tax revenues per taxpayer and (ii) the total revenues that are due to all other types of revenues expressed as a percentage of total revenues. This mismatch, maybe due to the fact that the effective tax rates that we calculate for each taxpayer are based on a simple and direct application of the tax brackets and marginal tax rates given in Table 8.2. However, this simplistic calculation does not account for the complexities in the tax code and thus does not consider deductions, credits, nor the amounts of tax refunds. In order to best translate our model outputs and interpret them at the national level, we chose to make a simple multiplicative transformation in order to reproduce the target individual income tax revenues and the overall debit for 2016. Thus, for each calibrated model run we chose to scale down the individual income tax revenues by a multiplicative factor.

### 9.4 The total costs to the GS

In 2016 the US government spent $3.863 trillion leading to a per capita cost of $11,923. We assume that the GS provides and maintains a fixed set of public services (i.e., schools, roads, policing, etc.) that have a total yearly fixed cost $C_{\text{Fixed}}$. The GS can change the yearly funds it provides the IRS to carry out its auditing activity and thus incurs a variable cost $C_{i}^{\text{Variable}}$. The GS can also change its fiscal policy by changing the tax brackets and the tax rates applied to each tax bracket. Changes in fiscal policy will have an effect on both taxpayer’s tax compliance behavior and the GS variable...
costs and projected revenues. As mentioned in Section 8.4 we used the 2016 IRS data tables to find that IRS audit-related expenditures of enforcement and collection were $2.329 billion. Thus, the per capita rate of this variable cost is just $7.2 per citizen and $15.7 per taxpayer. In our model, we assume that any unused funds allocated for IRS enforcement and collection are returned back to the GS at the end of the year.

The average cost per audit produced by our calibrated best case run of the model is $2,450 which corresponds to a total national level cost of $2.96 billion. This represents our status-quo variable cost. The fixed cost is thus obtained by subtracting $2.96 billion from the $3.863 trillion. This leads to a per capita fixed cost of $11,914. Therefore, in our model the fixed costs are given by

\[ C^{\text{Fixed}} = \alpha_C N, \quad (9.3) \]

where \( \alpha_C = 25,563 \) (i.e., equal to $11,914 \cdot 324/151) is the total fixed cost spending per taxpayer. For the status-quo conditions, the variable cost is very small and generally negligible. However, if the audit rate \( q \) is significantly increased, this cost could become important. For example, if the audit rate were to be significantly increased to say 12.5\%, the \( C_i^{\text{Variable}} \) could exceed 1\% of the total government spending.

### 9.5 The marginal budget deficit

When total revenues \( Q_t \) do not cover the total cost \( C_t = C^{\text{Fixed}} + C_t^{\text{Variable}} \), the GS will accrue a budget deficit \( C_t - Q_t \) which is added to the overall debt of the GS. In 2016 the national deficit and debt were respectively $586 billion and $19.59 trillion (i.e., $3,882 and $129,750 per taxpayer). The public debt, which includes all debt that is publicly owned and excludes all intra-governmental debt was $14,128. Since the deficit is greater than the tax gap, it is apparent that achieving full tax compliance would be insufficient at balancing the national budget without additional measures to either increase tax revenues or cut spending. Therefore, in our model, the total revenues can never cover the total costs. Therefore, we choose to track the marginal increase in the budget deficit and total debt with respect to the 2016 values. Our marginal deficit and debt are respectively denoted as \( \Lambda_t = C_t - Q_t -$586 \) billion, and \( D_{t+1} = D_t + \Lambda_t, \) where \( D_0 = 0. \)

### 9.6 The fiscal policy of the GS

The primary goal of the GS is to provide services to its citizens while minimizing the average yearly deficit over the long term. To achieve this goal, the GS aims to prevent and reduce the level of tax evasion, ensuring that the inflow of revenue is sufficient to maintain the functioning of services promised to taxpayers. It can achieve this goal by a combination of fiscal and tax evasion deterrence policies. The GS prefers to achieve its aim using a fiscal policy that keeps tax rates as low as possible. The GS could choose strategies that could significantly increase the average yearly deficit over the short term. For example, it could choose a combination of reducing tax rates and increasing funding to the IRS to fund its deterrence policy. It would choose to do this in the hope of increasing compliance and reducing the average yearly deficit over the long term. The GS adapts its strategy (i.e., fiscal and deterrence policies over time) based on the shortage of revenues collected by taxpayers as compared with the GS’s budget. The GS can also choose strategies that temporarily reduce taxation levels by running a yearly budget deficit in hopes of recovering a higher level of tax compliance in future years.

---

26 The variable cost is calculated using the cost table 8.3 and is based on the number of correspondence and field audit carried out by the IRS.
9.7 Estimating the changes in the US Government Debt to GDP ratio

In its attempt to recover a higher level of tax compliance, the GS will try to avoid reaching an unacceptable debt level which it would no longer be able to sustain. This unacceptable debt level is best measured as a percentage of the Gross Domestic Product (GDP). In 2015, the US had a GDP of $18.46 trillion and its debt to GDP was 105%. The question of what value should the debt be as a percentage of the GDP for the US is the subject of ongoing research. Some economists suggest that the total debt should be below 80% of GDP [68]. Others think that for large countries that are economically stable and growing the debt can be much larger [137]. Some countries that are viewed as economically stable have a very large debt to GDP. For example, according to the International Monetary Fund, Japan’s debt to GDP in 2016 was 238%. Nevertheless, although Japan is facing very serious problems with its debt, economists do not view Japan as being in imminent danger of a default as for example Greece that in 2016 had a lower debt to GDP of 160%. Moreover, in contrast to the US, most of Japan’s debt is owed internally to Japanese investors. Based on this, for our model, we set our indicative value for the critical or upper threshold value for the debt at 200% of GDP.

As mentioned previously, in our model achieving full tax compliance would be insufficient at balancing the national budget without additional measures to either increase tax revenues or cutting spending. However, if the growth rate of GDP continues to be high and interests payments on the debt stay relatively low then it is conceivable that there may be a fiscal and deterrence policy which would increase the level of tax compliance, and that it would allow the ratio of the debt to GDP to be maintained at its current level or even gently reduced over the years.

Our model tracks the total debt and estimates the changes in the ratio of debt to the GDP. The average growth rate of GDP in the US from 1990 to 2016 was 2.41% [27]. In our model, the growth in debt comes from the changes in the marginal deficit which takes into account the changes in individual income tax revenues. However, we also assume that as the economy grows, total revenues and most of the costs grow at the same rate. For example, we assume that the taxpayer’s income grows at the same rate as GDP. The only cost that is assumed to grow at a different rate is the amount that the federal government pays interest on the outstanding public debt. In 2016, this interest was $233 billion. This represents 1.65% of the public debt and 1.2% of the national debt. However, the interest rate varies from year to year at it is projected to strongly increase in the next decade [28]. Based on the above assumptions, the equation we use to provide a very approximate estimate of the dynamics of the debt dynamics is

\[ [\text{Debt}]_{t+1} = r_{\text{interest}} [\text{Debt}]_t + r_{\text{growth}}^t (\Lambda_t + 353 \cdot 10^9), \]

where \( t \) represents the number of years since 2016, \( [\text{Debt}]_0 = $19.59 \) trillion, \( r_{\text{interest}} = 1.012, \) \( r_{\text{growth}} = 1.0241 \) and $353 billion represents the deficit in 2016 when the costs of paying interest on the outstanding public debt is excluded. Using equation 9.4 our very approximate estimate of the ratio of debt to GDP is

\[ \text{Estimated debt to GDP ratio in year } t = \frac{[\text{Debt}]_t}{18.46r_{\text{growth}}^t}. \]

---


10 Design concepts

As shown in figure 6.1 our model describes tax compliance dynamics as a complex adaptive system (CAS) [125]. In Sections 6 and 7 we provided an overview of our ABM of tax compliance and described many elements of the CAS of tax compliance. Here, we collect and summarize the various leading concepts and components of the model that relates to the key features of CASs and describe what the model can be used for.

10.1 Adaptive and Interacting Agents

The main entity in our model are the taxpayers. Taxpayers have memory and they recall their past tax compliance decisions and outcomes. Their adaptive tax compliance behavior responds to (i) the taxes that they are asked to pay, (ii) their past outcomes with tax compliance, (iii) tax morale and risk perceptions of their alters in their social network and (iv) what they hear from the media. Taxpayers do not use a utility-maximization, forward-looking deductive thought that typically characterizes agents of a rational expectations model. Instead, taxpayers in our model are myopic and backward looking and are limited by bounded rationality. As such they are unable to correctly anticipate the expected individual and aggregated level outcomes. They instead turn to their social networks, use inductive thought and adapt to past outcomes to make their decisions.

The other entities in our model include the GS, the IRS, and the media. For the status-quo conditions, the GS applies a fixed fiscal and spending policy. Funds to the IRS are used to maintain the same audit rate and strategy every year. Thus, for the status-quo case, the GS and the IRS do not adapt. Media tracks the tax gap and past IRS activities and can interact with the agents by providing this information to the taxpayers.

10.2 Complexity, Emergence, and Criticality

Taxpayers are heterogeneous and in particular, have different initial risk perception and tax morale. Over time they self-segregate into taxpayers that are predominantly compliant and those that predominantly evade. The segregation into different behavioral groups is dynamic and taxpayers that are currently compliant can initiate or relapse back into tax evasion behavior, and those that are currently evading may become compliant due to the IRS deterrence activities. The formation of these different behavioral groups results from the adaptation process, social interactions and outcomes from the IRS deterrence activities. The collective decisions of the taxpayers affect the overall tax gap which in turn can affect the taxpayer’s compliance behaviors via media feedback loops. An additional coupling of the taxpayer’s compliance behaviors to the collective outcome can occur when the GS and the IRS are respectively allowed to change their fiscal and deterrence policy in response to the changing tax gap. Taxpayers’ adaptive tax compliance behavior together with changes in fiscal policies may lead to desirable or undesirable self-reinforcing dynamics. For example, this could lead to a vicious cycle with greater levels of evasion leading to larger taxation rates imposed by the GS on compliant taxpayers, consequently lowering perceived fairness even among those who initially perceived the system as fair.

10.3 Stocasticity

In our model, both the taxpayers and the IRS have chance driven (i.e., stochastic) effects and processes. The IRS uses its audit strategy to stochastically sample taxpayers to be audited for either a correspondence or a field audit. Taxpayers are influenced by various stochastic effects.
which determine: (i) whether they receive a tax refund or have a balance due, (ii) if and when they initiate in tax evasion behavior, (iii) the amount of income they choose to report after initiating in tax evasion behavior, (iv) whether they or their alters are selected for an audit and whether they are penalized, (v) whether they pay attention to the media reports about IRS audit activities.

10.4 Model purpose and use

Our model can be used to advance the understanding of the complex interplay between of tax fairness perceptions, risk perceptions, social networks, media, deterrence strategies and fiscal policies in producing the tax compliance aggregate behavior. Its purpose is to address (a) what are the determining assumptions and parameter values describing agents behavior that have the most leverage on the compliance levels, (b) what is the role of the initial compliance and network effects in determining the compliance level that the system will equilibrate towards, and (c) what are the conditions stability and for critical behavior in our system and which parameters determine critical points.

Our model can also be used by policy makers to gain better insights on what deterrence and fiscal policies work in combination that increase tax compliance and reduce the national debt. It can be used to find economically acceptable strategies that can help transform the culture of a population and reduce tax evasion. For example, the model can explore policies that start with an increase in audit rates and penalty rates and decrease taxation levels. This strategy would increase the national debt in the short run but might increase compliance in the long run. Eventually, taxation rates may have to increase to repay the debt, and if done at a slow enough rate, this strategy could ensure that compliant taxpayers were not lost in the process.

11 Population Heterogeneities and Initialization

In this section, we describe the attributes of our taxpayers of our ABM, their heterogeneities, and initial values. We also describe the data and methods used to inform the social network describing the relationships between taxpayers involving tax related interactions.

11.1 The taxpayer’s attributes

Table 11.1 provides a list of the dynamic attributes of the taxpayers used by our model and that are most relevant. Some dynamic attributes such as the under.reporting.history are numerical vectors of length \( K \) (i.e., the maximum number of past years that the IRS will investigate for unpaid taxes) and are used to calculate things like the total amount of back taxes and penalties to pay on unreported income. Other attributes are type logical and as such their value can be either true or false. For example, the attribute refund.return tracks whether the taxpayer received a tax refund in the previous year. These taxpayers attribute change endogenously according to the behavioral rules described in our ABM. Most of these attributes have been described in Sections 6.2 and 7.

Table 11.2 provides a list of all taxpayer’s static attributes. Most of these attributes are self-explanatory from the first column and have been described in Sections 6.2 and 7. Some other attributes are less clear and we next proceed to describe these.

We start with the proportion of the hideable income of the taxpayer (i.e., \( prop.hideable \)) which is the proportion of the taxpayer’s income that s/he could under-report to the IRS. The proportion of hideable income measures the level of opportunity that taxpayers have in terms of underreporting their income to the IRS. In our model, self-employed taxpayers have this attribute set to 100%. For
<table>
<thead>
<tr>
<th>Dynamic attribute</th>
<th>Symbol</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>tax.rate</td>
<td>$T_i^{(i)}$</td>
<td>Numerical</td>
<td>The effective tax rate.</td>
</tr>
<tr>
<td>V.new</td>
<td>$V_i^{(i)}$</td>
<td>Numerical</td>
<td>The pro-compliance experience count.</td>
</tr>
<tr>
<td>propensity.prop.income.report</td>
<td>$w_i^{(i)}$</td>
<td>Numerical</td>
<td>Probability to fully comply for compliant taxpayers.</td>
</tr>
<tr>
<td>per.audit.rate</td>
<td>$q_i^{(i)}$</td>
<td>Numerical</td>
<td>Perceived audit rate.</td>
</tr>
<tr>
<td>per.penalty.rate</td>
<td>$\bar{P}_i^{(i)}$</td>
<td>Numerical</td>
<td>Perceived penalty rate.</td>
</tr>
<tr>
<td>report.abs</td>
<td>$A_i^{(i)}$</td>
<td>Numerical</td>
<td>Dollar amount of income reported.</td>
</tr>
<tr>
<td>report</td>
<td>$a_i^{(i)}$</td>
<td>Numerical</td>
<td>Proportion of hideable income reported for non-compliant taxpayers.</td>
</tr>
<tr>
<td>w.Q</td>
<td>$T_i^{(i)}$</td>
<td>Numerical</td>
<td>Discounted count of audit experiences.</td>
</tr>
<tr>
<td>refund.return</td>
<td></td>
<td>Logical</td>
<td>Recently received a tax refund.</td>
</tr>
<tr>
<td>compliant</td>
<td></td>
<td>Logical</td>
<td>Recently compliant.</td>
</tr>
<tr>
<td>audited</td>
<td></td>
<td>Logical</td>
<td>Recently audited.</td>
</tr>
<tr>
<td>audited.non.compliant</td>
<td></td>
<td>Logical</td>
<td>Recently audited and not compliant.</td>
</tr>
<tr>
<td>penalized</td>
<td></td>
<td>Logical</td>
<td>Recently penalized.</td>
</tr>
<tr>
<td>penalty.history</td>
<td></td>
<td>Vector of length $K$</td>
<td>Penalty history in recent years.</td>
</tr>
<tr>
<td>audit.history</td>
<td></td>
<td>Vector of length $K$</td>
<td>Audit history in recent years.</td>
</tr>
<tr>
<td>under.reporting.history</td>
<td></td>
<td>Vector of length $K$</td>
<td>Reporting history in recent years.</td>
</tr>
</tbody>
</table>

Table 11.1: Taxpayer’s dynamic attributes of the taxpayers. The first column provides the name of the attribute used in the R code of the implementation of our model. The second column provides the mathematical symbol of the attribute that we used in our description of the model. The third column provides the data type of the attribute.

the non-self-employed taxpayer, the proportion of hideable was informed by IRS tables and is
<table>
<thead>
<tr>
<th>Static Attribute</th>
<th>Symbol</th>
<th>Type</th>
<th>Source</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>tax.ids</td>
<td></td>
<td>Numerical</td>
<td>Internal</td>
<td>Internal tax id</td>
</tr>
<tr>
<td>person.id</td>
<td></td>
<td>Numerical</td>
<td>NDSSL</td>
<td>NDSSL person id</td>
</tr>
<tr>
<td>household.id</td>
<td></td>
<td>Numerical</td>
<td>NDSSL</td>
<td>NDSSL household id</td>
</tr>
<tr>
<td>zipcode</td>
<td></td>
<td>Categorical</td>
<td>NDSSL</td>
<td>zipcode</td>
</tr>
<tr>
<td>age</td>
<td></td>
<td>Numerical</td>
<td>NDSSL</td>
<td>age</td>
</tr>
<tr>
<td>gender</td>
<td></td>
<td>Categorical</td>
<td>NDSSL</td>
<td>gender</td>
</tr>
<tr>
<td>household.size</td>
<td></td>
<td>Numerical</td>
<td>NDSSL</td>
<td>Number of people in the household.</td>
</tr>
<tr>
<td>num.dependents</td>
<td></td>
<td>Numerical</td>
<td>NDSSL</td>
<td>Number of dependents in the household.</td>
</tr>
<tr>
<td>income</td>
<td>$I^{(i)}$</td>
<td>Numerical</td>
<td>NDSSL &amp; Census</td>
<td>Household income</td>
</tr>
<tr>
<td>prop.hideable.income</td>
<td></td>
<td>Numerical</td>
<td>IRS Data</td>
<td>Proportion of the income that can be hidden from the IRS.</td>
</tr>
<tr>
<td>filing.status</td>
<td></td>
<td>Categorical</td>
<td>IRS Data</td>
<td>One of four filing statuses.</td>
</tr>
<tr>
<td>c1.tri.dist</td>
<td>$c_1$</td>
<td>Numerical</td>
<td>Expert Opinion</td>
<td>Values for $c_1^{(i)}$ based on a first distribution.</td>
</tr>
<tr>
<td>c1.alp.majority.fit</td>
<td>$c_1$</td>
<td>Numerical</td>
<td>ALP Survey</td>
<td>Values for $c_1^{(i)}$ based on a second distribution.</td>
</tr>
<tr>
<td>self.employed</td>
<td></td>
<td>Logical</td>
<td>ALP Survey</td>
<td>Self employed or waged worker</td>
</tr>
<tr>
<td>self.employed.propensity</td>
<td></td>
<td>Numerical</td>
<td>ALP Survey</td>
<td>Propensity to be self employed</td>
</tr>
<tr>
<td>per.audit.rate0</td>
<td>$\tilde{q}_{0}^{(i)}$</td>
<td>Numerical</td>
<td>ALP Survey</td>
<td>Baseline perceived audit rate.</td>
</tr>
<tr>
<td>per.penalty.rate0</td>
<td>$\tilde{P}_{0}^{(i)}$</td>
<td>Numerical</td>
<td>ALP Survey</td>
<td>Baseline perceived penalty rate.</td>
</tr>
<tr>
<td>actor.logical</td>
<td></td>
<td>Logical</td>
<td>ALP Survey</td>
<td>Exposed to the famous actor effect in the media.</td>
</tr>
</tbody>
</table>

Table 11.2: Taxpayer’s static attributes and sources. The fourth column provides the data source that informed the value or distribution of the attribute.

---

Table 11.3: The percentage of income that can be under-reported by adjusted gross income was informed by the IRS Table 1.4 for 2013 in the SOI Tax Stats - Individual Statistical Tables by Size of Adjusted Gross Income. For verification, we compared these numbers to those reported by Johns and Slemrod 2010 [84].

The filing status is a categorical attribute and a taxpayer can have one of the following four filing statuses: Single, Head of their Household, Married Filing Jointly or Married Filing Separately. The attributes c1.tri.dist.dist and c1.alp.majority.fit provide the range of c1 values which we sample from. Thus, as will be explained in Section 12.1 on model input parameters, the taxpayer’s actual value for c1 is stochastically chosen within this range. The attributes c1.tri.dist.dist and c1.alp.majority.fit are sampled from two distributions that are introduced later in this report where we describe the analyses of our survey. Specifically, these distributions are introduced in Section 14.2.4 and are shown in Figure 14.6. The attribute self.employed.propensity provides a probability that the taxpayer is self-employed. The attribute self.employed is static and consequently does not change over a simulation run. However, at the beginning of our simulation the self.employed attribute can be re-initialized and thus we can change who is self employed and who is not. This is done using the self.employed.propensity attribute and according to a stochastic selection of who is self-employed for the model run. This option allows us to carry out "what-if" policy scenarios to see what would happen if the proportion of self-employed taxpayers were decreased or increased. The actor.logical attribute was explained in Sections 5 and 7.4 and is related to the famous

<table>
<thead>
<tr>
<th>Income($k)</th>
<th>Percentage of hideable Income</th>
</tr>
</thead>
<tbody>
<tr>
<td>(0,5)</td>
<td>30.70%</td>
</tr>
<tr>
<td>[5,10)</td>
<td>20.70%</td>
</tr>
<tr>
<td>[10,15)</td>
<td>18.70%</td>
</tr>
<tr>
<td>[15,20)</td>
<td>11.30%</td>
</tr>
<tr>
<td>[20,25)</td>
<td>8.50%</td>
</tr>
<tr>
<td>[25,30)</td>
<td>7.30%</td>
</tr>
<tr>
<td>[30,40)</td>
<td>6.90%</td>
</tr>
<tr>
<td>[40,50)</td>
<td>6.70%</td>
</tr>
<tr>
<td>[50,75)</td>
<td>7.40%</td>
</tr>
<tr>
<td>[75,100)</td>
<td>8.70%</td>
</tr>
<tr>
<td>[100,200)</td>
<td>12.80%</td>
</tr>
<tr>
<td>[200,500)</td>
<td>28.30%</td>
</tr>
<tr>
<td>[500,1,000)</td>
<td>47.00%</td>
</tr>
<tr>
<td>[1,000,1,500)</td>
<td>57.50%</td>
</tr>
<tr>
<td>[1,500,2,000)</td>
<td>63.00%</td>
</tr>
<tr>
<td>[2,000,5,000)</td>
<td>68.70%</td>
</tr>
<tr>
<td>[5,000,10,000)</td>
<td>74.50%</td>
</tr>
<tr>
<td>10,000+</td>
<td>89.70%</td>
</tr>
</tbody>
</table>

of Adjusted Gross Income to find the percentage of income that can be under-reported. Using this table we considered the following sources of income that can potentially be hidden from tax authorities: 1) Taxable interest; 2) Tax-exempt interest; 3) Ordinary dividends; 4) Qualified dividends; 5) Business or profession net income minus Business or profession net loss; 6) Capital gain distributions reported on Form 1040; 7) Sales of capital assets reported on Form 1040, Schedule D [taxable net gain minus taxable net loss]; 8) Sales of property other than capital assets [net gain minus net loss]; 9) Total rental and royalty [net income minus net loss]; 10) Partnership and S corporation [net income minus net loss]; 11) Estate and trust [net income minus net loss]; 12) Farm [net income minus net loss]; 13) Foreign-earned income exclusion; 14) Other income [net income minus net loss]; 15) Gambling earnings; 16) Cancellation of debt.
person/actor effect.

The static attributes of the population of taxpayers that enter the model were informed by various sources. Our main source comes from the freely available data provided by the Network Dynamics and Simulation and Science Laboratory (NDSSL) Virginia Polytechnic Institute and State University which represents a synthetic population of the city of Portland, OR [114]. Sources from the Census and the IRS informed the income distribution and filing status. Behavioral attributes were informed by a survey of 1030 respondents that we fielded in 2016 on tax perceptions using the RAND American Life Panel (ALP) [30]. The survey informed taxpayer’s behaviors and network characteristics including (i) the average degree of social relationships that involve interactions about taxes, (ii) the level of assortative mixing by income and employment, (iii) perceptions of alters being audited. An overview of our survey questions and the statistical analyses that were used to inform model parameters and behavioral attributes is described starting from Section 13.

11.2 The Synthetic Population Network

The NDSSL data for Portland, OR provides an instance of a time-varying social contact network for a normative day, derived from daily activities [31]. The data was created from the urban transportation ABM to simulate the daily movements of individuals across locations in Portland, OR. From this, an edge list connecting nodes or vertices representing individuals was constructed [97]. The edge weights represent the duration of a typical day that individuals interact with each other. The original dataset contains ~ 1.6 million individuals, ~ 630 thousand households with an edge list of ~ 19.68 million face-to-face daily contacts across various types of activities and locations. For the purpose of our model, we chose to use this data to create a set of smaller networks by pruning the edge list and number of vertices (i.e., individuals). These smaller networks connect adults who are the heads of households through social activities and durations whereby income tax discussion is more likely to occur. The method used to prune the network does have limitations which we will discuss later in this section. Using the NDSSL household data, we obtained the household size, the number of dependents and, by analyzing the household structure and using IRS tables [32] we assigned one of four filing statuses to each household. We used both the NDSSL income category field and census data for household income [12] to generate the incomes for each taxpayer [33]. The household income and number of dependents in the household were split stochastically between married couples that file separate tax returns.

The process of reducing the network to one containing just working adults who are heads of their households and pruning the connections by activity and duration produced a network with ~ 460 thousand taxpayers and ~ 7 million face-to-face daily contacts, with a mean degree of 30.37 alters per taxpayer [34]. We refer to this synthetic population and network as PN460 (Portland Network 460). However, this network is still very large to use to run many different cases of our model, limiting our ability to use this network for verification and validation of the ABM.

---

30 The ALP is a nationally representative, a probability-based panel of over 6000 members ages 18 and older who are regularly interviewed over the internet for research purposes. ALP survey data sets can be obtained at https://alpdata.rand.org/
31 The NDSSL data for Portland, OR is available at http://ndssl.vbi.vt.edu/synthetic-data/download.html
32 Number of Returns, by Age, Marital Status, and Size of Adjusted Gross Income, the tax year 2014 is available at https://www.irs.gov/uac/soi-tax-stats-individual-statistical-tables-by-filing-status
33 We used the Current Population Survey (CPS) of 2015 data for Family Income Table that is available at https://www.census.gov/data/tables/2015/demo/cps/ui-01.html
34 For our purpose, we created an edge list for this network that is unweighted and where connections represents a social interaction between two individuals for a significant duration of the day and over an activity or interaction that suggests that the interaction is between individuals who know each other.
Therefore, we chose to prune the network further to produce smaller networks that still retain the important statistical characteristics of PN460 and that can be more easily used by our ABM to run many model cases. We generated two networks with roughly one thousand and ten thousand taxpayers which we refer to as PN1 and PN10 respectively. We reviewed the literature on the recommended ways to sample from a network and produce smaller representative networks of the original. We chose to use a simplistic sampling approach that essentially selected a small set of closely separated initial nodes (e.g., 3 or 4) and their first, second and sometimes third degree contacts on the PN460 network. Depending on the starting nodes and on how many degrees of separation we chose to sample to, the resulting network contained roughly 1 thousand and 10 thousand nodes. In this process, the selected nodes from the last degree of separation (i.e., peripheral nodes) have edges connecting them to a higher degree of separation nodes that are beyond our sample. We call these dangling edges or bonds, and these need to be rewired to the nodes belonging to our sample in such a way that certain mixing statistics are still representative of the original PN460 network. The rewiring method we chose to use preserves mixing distribution by income and by zip code and connected the dangling edges between peripheral nodes in a way that best preserved the degree distribution of the original network. Figure 11.1 shows the degree and location distributions for networks PN460, PN10 and PN1. Other, network statistics we used to make comparisons between the networks were the income and zip code mixing matrices. Therewiring method we chose to use preserves mixing distribution by income and by zip code and connected the dangling edges between peripheral nodes in a way that best preserved the degree distribution of the original network. Figure 11.1 shows the degree and location distributions for networks PN460, PN10 and PN1. Other, network statistics we used to make comparisons between the networks were the income and zip code mixing matrices. Table 11.6 shows the quantiles of the income distribution for PN10. The income distributions for the PN1 and PN460 were also similar. The income distribution was informed by both the NDSSL data for Portland, OR and 2015 Census data on income and was adjusted such that the mean income of our taxpayers was $71,000. This value was chosen to roughly agree with the $70,861 calculated and described in Section 9.1. The income information from the NDSSL data was used to identify the taxpayer ranking in terms of income from which we sampled from the adjusted 2015 Census data for income.

<table>
<thead>
<tr>
<th></th>
<th>18-25 (%)</th>
<th>25-35 (%)</th>
<th>36-50 (%)</th>
<th>51-64 (%)</th>
<th>65+ (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PN1</td>
<td>7.03</td>
<td>30.21</td>
<td>41.42</td>
<td>18.66</td>
<td>2.68</td>
</tr>
<tr>
<td>PN10</td>
<td>6.97</td>
<td>27.66</td>
<td>43.97</td>
<td>18.99</td>
<td>2.41</td>
</tr>
<tr>
<td>PN460</td>
<td>4.17</td>
<td>24.83</td>
<td>44.52</td>
<td>20.30</td>
<td>6.18</td>
</tr>
</tbody>
</table>

Table 11.4: Percentages showing age groups of our tax filers for each network.

As can be seen from Figure 11.1 and Tables 11.4 and 11.5, this network pruning has very apparent and important limitation. These include (i) it under-samples nodes with very few edges and does not sample isolates who have no contacts at all, and (ii) the sampling is biased on the how the initial nodes are selected. So although the overall income distribution and mixing are not significantly affected, other distributions such as age and gender attributes are less representative of the original PN460 network and of US tax-payers. Such limitations will be addressed in future research.

The edges in our network represent social contacts between the head of households and tax filers. The average degree of our network is 30.37. Social contacts in this network represents face-to-face daily contacts among people who are likely to know each other. These include con-

\[\text{Zip codes were obtained from using the location data in the NDSSL data for Portland, OR and finding the zip codes using the GIS data available at } \text{http://gis-pdx.opendata.arcgis.com/datasets/8efb5a8857da4a7094ed2348a94f5789_1} \]

\[\text{The mean income across filing status presented in table } 11.6 \text{ can be compared to those provided at } \text{https://www.fool.com/retirement/2017/03/04/whats-the-average-americans-tax-rate.aspx and } \text{https://www.fool.com/retirement/2016/10/30/heres-the-average-american-household-income-how-do.aspx}\]
Figure 11.1: Plots showing the degree distribution (column 1) and the income distribution (column 2) of the household location for Portland, OR using networks (A) PN460, (B) PN10 and (C) PN1.

<table>
<thead>
<tr>
<th></th>
<th>Head of Household (%)</th>
<th>Married Filing Jointly (%)</th>
<th>Married Filing Separately (%)</th>
<th>Single (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PN1 Females</td>
<td>6.03</td>
<td>6.53</td>
<td>1.76</td>
<td>11.28</td>
</tr>
<tr>
<td>PN1 Males</td>
<td>3.18</td>
<td>57.91</td>
<td>2.51</td>
<td>10.81</td>
</tr>
<tr>
<td>PN10 Females</td>
<td>5.80</td>
<td>7.81</td>
<td>1.70</td>
<td>11.09</td>
</tr>
<tr>
<td>PN10 Males</td>
<td>3.27</td>
<td>57.54</td>
<td>2.31</td>
<td>10.48</td>
</tr>
<tr>
<td>PN460 Females</td>
<td>7.54</td>
<td>7.01</td>
<td>3.93</td>
<td>8.05</td>
</tr>
<tr>
<td>PN460 Males</td>
<td>3.76</td>
<td>56.95</td>
<td>4.17</td>
<td>8.61</td>
</tr>
</tbody>
</table>

Table 11.5: Percentages showing gender and filing status of our tax filers for each network.

tacts amongst people who know each other relatively well such as amongst family members, co-workers, and friends. They also include contacts with people who know each other more casually and for other reasons. For example, people that provide a service such as nannies, teachers, doc-
Table 11.6: Quantiles of the income distribution of the PN10 network.

<table>
<thead>
<tr>
<th>Filing Status</th>
<th>Min.</th>
<th>1st Qu.</th>
<th>Median</th>
<th>Mean</th>
<th>sd</th>
<th>3rd Qu.</th>
<th>Max.</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>PN10</td>
<td>$5</td>
<td>$24,950</td>
<td>$50,827</td>
<td>$71,000</td>
<td>$84,592</td>
<td>$89,587</td>
<td>$1,015,257</td>
<td>10112</td>
</tr>
<tr>
<td>Head of Household</td>
<td>$5</td>
<td>$14,295</td>
<td>$27,320</td>
<td>$35,857</td>
<td>$40,048</td>
<td>$49,048</td>
<td>$183,043</td>
<td>917</td>
</tr>
<tr>
<td>Married Filing Jointly</td>
<td>$5</td>
<td>$35,819</td>
<td>$67,780</td>
<td>$88,543</td>
<td>$95,476</td>
<td>$109,776</td>
<td>$1,015,257</td>
<td>6608</td>
</tr>
<tr>
<td>Married Filing Separately</td>
<td>$37</td>
<td>$11,737</td>
<td>$19,676</td>
<td>$54,392</td>
<td>$99,605</td>
<td>$58,986</td>
<td>$942,111</td>
<td>406</td>
</tr>
</tbody>
</table>

Figure 11.2: Illustration of the PN1 Network. Colors represent income categories from dark red to navy blue in group breaks at $25,000, $50,000, $100,000, $250,000 and $500,000. The nodes with the diamond symbol are self-employed. The position of the nodes was calculated based on income, age group, and self-employment status. The edges represent social relationships that could lead to interactions about taxes.

This latter group of people are unlikely to be considered close contacts and even less likely to engage in tax-related discussions. Therefore, in the context of tax evasion behavior, a very small fraction of these social contacts (i.e., alters) lead to interactions and information spreading about tax morale and tax risk perceptions. In our ALP survey, which is described starting from Section 13, we obtained information of the alters of each respondent. However, respondents were limited in eliciting a maximum of ten alters with which they have relationships and interact the most. Respondents were asked a variety of questions regarding their alters including whether or not they talking about taxes and if so how frequently. Respondents were also asked whether they
and their alters are currently working and if so whether they are self-employed or working for a wage/salary. We used this information to assign an attribute to each of our network edges which describe whether the edge is active or not in acting as a social link where information and perception of tax compliance behavior can spread. The sampling of active edges is done in order to (i) maintain the average degree of active connections in our taxpayer population and (ii) maintain the mixing matrix describing the proportion of edges connecting self-employed taxpayers with each other and with those who work for a wage. The self-employment attribute was assigned using ALP survey data and a regression model that is described in Section 15. Table 11.7 provides a summary of the results of the survey that were used to inform how we sample active contacts. Details on how these numbers were found from the survey are explained in Section 15. Out of

<table>
<thead>
<tr>
<th>Degree or prop. effect on the degree</th>
<th>value</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean degree</td>
<td>9.33</td>
</tr>
<tr>
<td>mean degree talk taxes</td>
<td>2.63</td>
</tr>
<tr>
<td>Lower bound (LB) degree talk taxes</td>
<td>2.00</td>
</tr>
<tr>
<td>Upper bound (UB) degree talk taxes</td>
<td>4.00</td>
</tr>
<tr>
<td>prop. SE-SE</td>
<td>0.44</td>
</tr>
<tr>
<td>prop. SE-NSE</td>
<td>0.30</td>
</tr>
<tr>
<td>prop. NSE-SE</td>
<td>0.28</td>
</tr>
<tr>
<td>prop. NSE-NSE</td>
<td>0.25</td>
</tr>
<tr>
<td>audit multiplicative effect</td>
<td>1.28</td>
</tr>
</tbody>
</table>

Table 11.7: Table showing inputs to informing the contact network of tax morale and tax risk perceptions obtained from the ALP survey. In this tables SE stands for self-employed and NSE for not-self-employed.

a maximum of 10 alters that respondents in our survey could elicit, on average our panel listed 9.33 alters with whom they have social contacts, and of those contacts, they talk to 2.63 alters per year about taxes. By analyzing the quantiles of the distribution of contacts, we assumed that this average could range between 2 and 4. However, as described in more detail in Section 15, this average depends on the type of work of the respondent and of his/her alters. We found that 44% of the contacts or edges between self-employed respondent and a self-employed alters also involve discussions of taxes. Instead, 25% of the contacts or edges between not-self-employed respondent and a not-self-employed alters also involve discussions of taxes. We further found that respondents who were audited in the past had about 1.28 times more alters with whom they spoke about taxes than those who were never audited. We used the numbers provided in Table 11.7 for our sampling of the active edges in our network. This produced networks where edges were active if they involved talking about taxes, leading to an average degree of 2.63 alters per taxpayer.

Our model further allows for the active/un-active attribute of each edge to change dynamically. Stochastically, every taxpayer in our model can change all of their active edges every tax generation period $\tau_G$ and resample them from his/her full set of edges. This process changes his/her degree of active contacts. It can of course also change the degree of other taxpayers affected by the resampling of the edges. Furthermore, if a taxpayer is audited, s/he will increase her degree by sampling additional edges to activate according to the 1.28 multiplicative factor.

## 12 Input data

Our ABM was informed by both primary and secondary data sets. In this section, we describe the various datasets and their sources, and how they were used to inform our ABM. We focus on
the freely available secondary datasets we used. Then in Section 13, we will describe our primary data, namely the survey data collected via the American Life Panel (ALP).

12.1 Model parameter values and related data sources

We start by listing and categorizing all model parameters. Then, for each parameter, we list the range of values that it can take in our model and describe the data or the literature we used to make our estimates. Tables 12.1 to 12.3 provide a list of the model parameters described and introduced in Sections 6 and 7. The table provides (i) the parameter names as they appear in the model implementation code, (ii) the mathematical symbol used, (iii) the type or group which the model parameter belongs to, (iv) the data source used to inform the parameter value and (v) a brief description of the model parameter.

Estimated parameter values and their ranges are shown in tables 12.4, 12.5 and 12.6. The reason we specify a range of values for each parameter is so that we can run (i) an uncertainty analysis to describe the range of possible model outcomes given a set of model parameter input values and (ii) a sensitivity analysis to determine how sensitive model outputs are to variation of individual model input parameter values and thus identify and rank parameters based on their importance in the prediction process of future cases. We created a large ensemble of plausible model realizations (which we also refer to as case runs) based on the systematic variation of the values of our parameters within the specified ranges. Each realization considers a unique combination of parameter values sampled using an appropriated probability distribution (e.g., uniform or a beta distribution) from the uncertainty ranges. We used a Latin Hypercube Sampling (LHS) process to select our unique set of parameter values combinations [77]. Therefore, the last column in tables 12.4, 12.5 and 12.6 indicates the sampling distribution used by our LHS. Uniform indicates a continuous uniform distribution. Discrete uniform indicates that the parameter takes integer values and is sampled with within its range with equal probability. PERT indicates that the parameter is sampled from a Program Evaluation and Review Technique distribution. The PERT distribution is a special type of beta distribution and is described in Appendix C. However, some parameters that appear in tables 12.5 and 12.6 are not independently sampled between the values specified in the range. Instead, they directly depend on the value that was sampled for another parameter. For these parameters, the sampling distribution is labeled as 'Derived'. For example, in table 12.6 the sampled value of the $\alpha_{GF}$ is derived from the value for $\gamma_{GF}$. Thus, when the value for $\gamma_{GF}$ takes its maximum value of 0.183, the value for $\alpha_{GF}$ also automatically takes its maximum value of 0.980. Thus, they are perfectly correlated. If there is an inverse dependence then the sampling is such that the two parameters are perfectly inversely correlated. So for example, when the value for $\beta_P$ takes its minimum value of 0.45, the values for $\beta_N$ and $\beta_M$ automatically take their maximum values of 0.275, and thus the sum $\beta_P + \beta_N + \beta_M$ is always equal to one.

All but one of the model parameters contained in tables 12.1 to 12.3 have been introduced in previous sections that describe our model. Here, we describe the parameter $c1.dist.weight$ which has not been previously introduced. As explained in the Section 11.1, each taxpayer has two attributes $c1.tri.dist.dist$ and $c1.alp.majority.fit$ which provides the bounds of his/her range of values for $c1$. In other words, his/her $c1$ value lies between these two bounds. For a given case run of our ABM, the $c1$ value of a taxpayer is sampled using a weighted average of his/her $c1.tri.dist.dist$ and $c1.alp.majority.fit$. The parameter $c1.dist.weight$, provides the weight that we use to assign the taxpayers $c1$ values. Therefore, by changing the value of $c1.dist.weight$ within [0,1] we can consider different cases where the distribution of the taxpayers $c1$ is between the two distributions introduced in Section 14.2.4 and shown in Figure 14.6.
<table>
<thead>
<tr>
<th>Name</th>
<th>Symbol</th>
<th>Type</th>
<th>Source</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>audit.rate</td>
<td>$q_t$</td>
<td>Controls</td>
<td>IRS data</td>
<td>Audit rate</td>
</tr>
<tr>
<td>detection.eff</td>
<td>$\epsilon_A$</td>
<td>Controls</td>
<td>IRS data</td>
<td>Detection Efficiency</td>
</tr>
<tr>
<td>K</td>
<td>$K$</td>
<td>Controls</td>
<td>Expert Opinion</td>
<td>Maximum years of back audits</td>
</tr>
<tr>
<td>penalty.rate</td>
<td>$p_t$</td>
<td>Controls</td>
<td>IRS data</td>
<td>Penalty rate</td>
</tr>
<tr>
<td>rate.refund.movement</td>
<td>$\rho$</td>
<td>Refund</td>
<td>Expert Opinion</td>
<td>Overall transition rate for tax refund status</td>
</tr>
<tr>
<td>return.weight</td>
<td>$r_w$</td>
<td>Refund</td>
<td>Literature</td>
<td>Weight placed on receiving a tax refund on the compliance behavior</td>
</tr>
<tr>
<td>bomb.crater.factor</td>
<td>$\kappa_{BC}$</td>
<td>Deterrence</td>
<td>ALP Survey</td>
<td>Multiplicative coefficient applied on the perceived audit rate describing the bomb crater effect.</td>
</tr>
<tr>
<td>gamblers.fallacy.grad</td>
<td>$\gamma_{GF}$</td>
<td>Deterrence</td>
<td>ALP Survey</td>
<td>First parameter that determines the multiplicative coefficient applied on the perceived audit rate describing the gambler’s fallact.</td>
</tr>
<tr>
<td>gamblers.fallacy.intercept</td>
<td>$\alpha_{GF}$</td>
<td>Deterrence</td>
<td>ALP Survey</td>
<td>Second parameter that determines the multiplicative coefficient applied on the perceived audit rate describing the gambler’s fallact.</td>
</tr>
<tr>
<td>m.qP</td>
<td>$m_x$</td>
<td>Deterrence</td>
<td>ALP Survey</td>
<td>Percieved deterrence value where a taxpayer’s $c_1^{(i)}$ value is half way between his/her $c_1^{(i)}$ and $c_2$.</td>
</tr>
<tr>
<td>s.qP</td>
<td>$s_x$</td>
<td>Deterrence</td>
<td>ALP Survey</td>
<td>Parameter controlling how fast a taxpayer’s compliance behavior changes with increasing perceived deterrence.</td>
</tr>
</tbody>
</table>

Table 12.1: Summary list of model parameters. The table provides a description of the type of parameter, in which section of the code it is mainly used and the data source used to inform it.

12.2 Control and Tax Refund Parameters

The parameters in table 12.4 describe rates for the dynamics of how many taxpayers receive a tax refund and the parameters that control and describe IRS audit activities. The parameters describing the audit rate, detection efficiency, and penalty rates were obtained by looking at IRS tables.
<table>
<thead>
<tr>
<th>Name</th>
<th>Symbol</th>
<th>Type</th>
<th>Source</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>beta.personal</td>
<td>$\beta_P$</td>
<td>Personal</td>
<td>ALP Survey</td>
<td>Weight place on personal-level evaluations on the compliance behavior</td>
</tr>
<tr>
<td>v.PP</td>
<td>$\theta$</td>
<td>Personal</td>
<td>Literature</td>
<td>Minimum value of the propensity $w_{i+1}^{(j)}$ for a taxpayer that has being penalized in the previous year.</td>
</tr>
<tr>
<td>c2</td>
<td>$c_2$</td>
<td>Personal</td>
<td>Literature</td>
<td>Maximum tax rate above which all taxpayers would underreport on their taxes.</td>
</tr>
<tr>
<td>c1.dist.weight</td>
<td>NA</td>
<td>Personal</td>
<td>Full Range</td>
<td>Sampling parameter that provides the weight placed on the first distribution for the $c_1$ values.</td>
</tr>
<tr>
<td>ave.degree.tTaxes</td>
<td>$\langle k_N \rangle$</td>
<td>Social</td>
<td>ALP Survey</td>
<td>Average number of tax-related social ties per taxpayer in the network</td>
</tr>
<tr>
<td>beta.network</td>
<td>$\beta_N$</td>
<td>Social</td>
<td>ALP Survey</td>
<td>Weight place on social network-level evaluations on the compliance behavior</td>
</tr>
<tr>
<td>beta.media</td>
<td>$\beta_M$</td>
<td>Media</td>
<td>ALP Survey</td>
<td>Weight place on media-level evaluations on the compliance behavior</td>
</tr>
<tr>
<td>media.mid.effect</td>
<td>$m_m$</td>
<td>Media</td>
<td>ALP Survey</td>
<td>The level of the taxgap reported by the media that would reduce the taxpayer’s media-level evaluations on compliance to 50%.</td>
</tr>
<tr>
<td>media.steepness</td>
<td>$s_m$</td>
<td>Media</td>
<td>ALP Survey</td>
<td>Parameter controlling how fast a taxpayer’s media-level evaluations on compliance changes with increasing taxgap.</td>
</tr>
<tr>
<td>tax.gap.reporting.media.threshold</td>
<td>$\Theta_M$</td>
<td>Media</td>
<td>Estimate</td>
<td>Minimum taxgap before the media begins to report on the problem of tax evasion.</td>
</tr>
</tbody>
</table>

Table 12.2: Summary list of model parameters continued.
<table>
<thead>
<tr>
<th>Name</th>
<th>Symbol</th>
<th>Type</th>
<th>Source</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>t.media.range</td>
<td>τ(\text{Media})</td>
<td>Time Scales</td>
<td>Estimate</td>
<td>Time scale considered by the media in its activity in reporting IRS deterrence rates.</td>
</tr>
<tr>
<td>generation.half.life</td>
<td>τ_G</td>
<td>Time Scales</td>
<td>ALP Survey</td>
<td>Expected number of years that a taxpayer pays taxes in his/her lifetime.</td>
</tr>
<tr>
<td>morale.half.life</td>
<td>τ</td>
<td>Time Scales</td>
<td>ALP Survey</td>
<td>Time scale controlling the discount parameter of tax related experiences of past years.</td>
</tr>
</tbody>
</table>

Table 12.3: Summary list of model parameters continued.

The average audit rate between years 2006 and 2015 was 1% and ranged between 0.84% in 2015 and 1.11% in 2010.\footnote{IRS audit data is available at \url{https://www.irs.gov/uac/soi-tax-stats-examination-coverage-individual-income-tax-returns-examined-irs-data-book-table-9b}.} We chose to set our mode value for the audit rate at 1% but sample it's value using a PERT distribution using a wide range. The choice of using a wide range allows us to better understand the sensitivity of the model to the audit rate while still using reasonable numbers. Note that with a PERT sampling distribution the minimum and maximum values of the range are hardly ever sampled. Our estimate for the mode value for the audit detection efficiency or effectiveness was 80%. This was based on Erard and Feinstein\footnote{The IRS Data Book 2014 can be found at \url{https://www.irs.gov/pub/irs-soi/14databk.pdf} and the IRS Penalty Interest Calculator.} and expert consultation.

As can be seen in Table 4 of Johns and Slemrod, 2010\footnote{IRS audit data is available at \url{https://www.irs.gov/uac/soi-tax-stats-examination-coverage-individual-income-tax-returns-examined-irs-data-book-table-9b}.}, the detection efficiency does not seem to depend on income. This is especially true for the self-employed (i.e., those with Schedule C income that are not subject to 3rd party information reporting). Detection efficiency increases depending on the extent of available third-party reporting. Generally, detection efficiency is larger for correspondence audits than it is for field audits. The range of values we chose for the efficiency reflect the possible efficiency range between correspondence and field audits. Moreover, correspondence audits usually check whether a taxpayer was compliant for the most recent single tax year whereas field audits may include a number of back audit years which could range between one to five. Roughly, correspondence audits account for 75% of all audits while field audits account for the other 25% and typically target higher-income taxpayers with more complicated returns. Based on expert opinion, we sample the value for the parameter describing the average number of back audits $K$ uniformly within this range. This parameter only applied to field audit and the actual number of back audits during a field audit of a taxpayer is stochastically sampled using a zero-truncated truncated Poisson distribution with mean $K$. Correspondence audits only involve an audit of the most recently filed tax returns.

Our estimate of the range of the penalty rate is based on considering accuracy-related penalties. We assume that any amount of income that a taxpayer desires to under-report is not disclosed as part of their tax return. We further assume that all taxpayer including those that under-report their income pay their due taxes to the IRS on time. We used the IRS Data Book for 2014 and the IRS Penalty and Interest Calculator\footnote{The IRS Data Book 2014 can be found at \url{https://www.irs.gov/pub/irs-soi/14databk.pdf} and the IRS Penalty Interest Calculator.}.
Table 12.4: Table showing IRS related parameters and other tax related tuning parameters.

In Sections 7.1 and 8.2 we described how we model tax refunds and their effect and described the parameters $r_w$ and $\rho$. Taxpayers that receive a tax refund are very likely to continue to receive a tax refund in the following year. Based on the expert opinion the expected time that a taxpayer could potentially transition between the two tax refund states is 10 years and it can range between 5 and 20 years. Thus, our mode value for $\rho$ was set to 0.1 per year and ranged between 0.05 and 0.2 per year. As explained in Section 8.2, the actual probability of transition per year is determined by the rate $\rho$ and a transition probability matrix with values that depend on the income of the taxpayer. The behavioral effect on compliance by a taxpayer receiving a tax refund is described by the parameter $r_w$ introduced in Section 7.1. The value for $r_w$ was informed from the results presented in columns two and three of Table 6 in Christian 1994 [39]. Column 2 presents the income understating percentages of those who do and do not receive a tax refund. We find that an estimate for the maximum effect of receiving a tax refund on compliance is 37\% (i.e., 1-(1-0.58)/(1-0.326)). Column 3 presents the voluntary compliance levels for these taxpayers. Using this column we find a smaller effect of 8\% (i.e., 1-(0.888)/(0.968)). We also had a few questions in our ALP survey to estimate this effect. This will be described in Section 13. Our survey found that tax refunds had no effect on tax compliance. Consequently, our lower bound for the value of $r_w$ is 0\%.

As explained in Section 7.1 the parameter value for $c_2$ represents the exasperating tax rate that taxpayers consider being completely unfair leading to a propensity to fully evade or stop working all together and not produce income. Our estimate for the value of $c_2$ is based on the estimates of the Laffer curve and in particular the estimated tax rate on the curve which would lead to no tax revenues. The Laffer curve has been estimated for the US and other countries [16, 29, 166, 167]. Estimates for $c_2$ based on the Laffer curve literature range between 65\% and 90\%. The mode value for $c_2$ we chose to use was based on the liability-based and payment-based linear models developed by Yu Hsing 1996 [29]. These estimates lead to a $c_2$ value between 65 and 68\%. We chose to round this to 70\%.

Consider the scenario whereby a taxpayer filed their tax return but instead of reporting the amount of taxes $A$ that they owe the IRS, they report a substantially lower dollar amount $A_R$. If the IRS audits and catches this taxpayer, s/he will have to pay the remaining $A - A_R$ and will be charged an accuracy-related penalty rate of 20\% plus interest on this amount. If instead the taxpayer under-reported their entire income (i.e., $A_R = 0$) and does not file a tax return and is audited and caught, will have to pay both the failure to file penalty and any accuracy-related penalties that sum to 45\% plus interest on $A$ dollars. The daily compound interest rate is assumed to be 4\% per year. Thus, by denoting $R_C = (1 + 0.04/365)^{365}$, the interest rate on taxes that are due $y$ tax years ago is given by $R_C^{0.5^y} - 1$. However, assuming the back audit is three years we chose to simplify and assume an average rate of 5\% that applies to all years. Therefore, our estimated penalty rate is 50\%. The lower bound value is 25\%. Our upper bound estimate is instead estimated as 80\% which is roughly the ratio between the total assessed civil penalties of $25.6$ billion and the total additional tax assessments of $33.1$ billion according to estimated by the IRS for 2014 ( IRS Data Book 2014 pages 24 & 44).
### 12.3 Time Scale Parameters

Parameters listed in table 12.5 refer to the timescales that enter the model. The first two parameters, $\tau$ and $\vartheta$ are key behavioral parameters that together control how many years a recently penalized taxpayer will it take before relapsing back into tax-evasion behavior and under-reporting most or all of his/her hideable income. Appendix [A] Section [A.3] analyses and explores the time scales of the EWMA process using different combinations of values of these three parameters and presents a table that helps inform the values we have chosen.

<table>
<thead>
<tr>
<th>Name</th>
<th>Symbol</th>
<th>Value</th>
<th>Range</th>
<th>Sampling Distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>morale.half.life</td>
<td>$\tau$</td>
<td>2</td>
<td>[1,8]</td>
<td>PERT</td>
</tr>
<tr>
<td>v.PP</td>
<td>$\vartheta$</td>
<td>0.90</td>
<td>[0.50,1.00]</td>
<td>PERT</td>
</tr>
<tr>
<td>generation.half.life</td>
<td>$\tau_G$</td>
<td>25</td>
<td>[10,35]</td>
<td>Derived from $\tau$</td>
</tr>
<tr>
<td>t.media.range</td>
<td>$\tau^{(\text{Media})}$</td>
<td>25</td>
<td>[10,35]</td>
<td>Derived from $\tau$</td>
</tr>
</tbody>
</table>

Table 12.5: Model parameters that control the various time scales involved in the model.

We start by describing the most important of the three parameters. As described in Section [6.2], the value of the half-life parameter $\tau$ determines the discount parameter $s$ used in the EWMA. With the exception of the experience of being penalized by the IRS, the parameter $\tau$ determines for how long past tax-related experiences and evaluations affect present decisions based on discounting past experiences exponentially. Modeling human discounting of past experiences when making decisions by using an EWMA approach has been investigated and applied in the fields of psychology, economics, and health. Examples include health prevention behavior [36, 37, 135, 171, 185], gambling behavior [48], finance investment behavior [30, 126] and driving behavior [63]. Frederick et al 2002 [58] provides a critical review of human time discounting and time preference using both an exponential and hyperbolic discounting function. Table 1 and figure 2 in the review article shows that there is a great amount of variability in the estimation of the annual discount factor and that most studies indicate that humans strongly discount past experiences. More recent studies produced similar findings [152, 162]. Moreover, as noted by Loewenstein [109, 110] and by Chapman [36], time-preference are delineated by "hot" and "cold" emotional states. Hot states include impulsive behaviors (e.g., gambling decisions) whereas cold states are more habitual behaviors (e.g., vaccination and cancer screening decisions). Their findings show that humans discount more steeply when they are in a hot emotional state. Thus, the appropriate value to use for the annual discount parameter depends on the different contexts and in particular on the frequency that people make decisions and evaluations in a repeated game. Tax compliance decisions are made yearly and can be made as either an impulsive reaction (e.g., during initiation) and from habitual consideration (e.g., a continued action). Therefore, we focused on the literature of repeated games with similar timescales. These include health prevention decisions. One example is health prevention behavior and in particular, the yearly decision to vaccinate for influenza [36, 37, 135, 171, 185]. In this literature, the typical half-life timescales quoted ranges from 1.7 to 3.6. Our choice for the range of values of $\tau$ shown in table 12.5 encompass these values and consider an even larger range. However, we sample this range using a PERT distribution with a mode value of 2. This mode value is taken from our previous work on modeling influenza vaccination decisions [171, 172].

The half-life and it’s derived discount parameter is used in the model to discount tax morale and risk perceptions. However, taxpayers that are caught for tax evasion will more strongly diminish the importance of the experiences in the years prior to being penalized. The strength of this discounting will depend on the amount of the penalty and of the amount of unpaid past taxes.
that are due. This discounting was described in Section 7.2 and is also described in Appendix A.3. There we introduced the tuning parameter \( \vartheta \). This parameter describes the effect on the tax morale component of a taxpayer that fully evades on taxes and is caught and penalized by the IRS. In such a case, the tax morale \( \Delta_{i}^{(t)} \) of taxpayer \( i \) increases to a value \( \vartheta \). A manual sensitivity analysis revealed that the model is not very sensitive to this parameter as long as it is reasonably high (e.g., \( \geq 60\% \)). We chose a large range of this parameter value which spans values between 0.5 and 1.0. The mode value of the PERT distribution that we chose is 0.9. This estimate was loosely informed by considering results published in the 2016 National Taxpayer Advocate (NTA) Audit Impact Study report [153] that is summarized in Beer et al. [20]. The report found that on average audited taxpayers that were caught and penalized for tax evasion increased their subsequent reported taxable income by 250\% percent after the audit. However, three years later the post-audit income reported increase reduced to only 120\% percent compared to the pre-audit reported income. Based on these numbers we estimate that taxpayers who under-reported all of their hideable income before an audit would report just 63\% of their hideable income three years later. By repeated manual approach of running our model for different values of \( \vartheta \), we found that values for \( \vartheta \) that are close to 0.9 were most consistent with the findings of the NTA report and the 63\% value.

The generation half-life \( \tau_{G} \) provides the typical time scale for the duration that a taxpayer works for and produces taxable income in a lifetime. So assuming that working taxpayers enter the workforce between ages 15 and 25 and exit between ages 55 and 75, the half-life of this time scale should range between 15 and 30 years (i.e., half of the number of years between 25 and 55, and between 15 and 75). As shown in the last column of Table 12.5 the value of \( \tau_{G} \) is "Derived from \( \tau \)" and thus is perfectly correlated with the sampled value for the parameter \( \tau \). Therefore, the sampling distribution for \( \tau_{G} \) is also a PERT\(^{40}\).

The last parameter listed in table 12.5 is \( \tau_{G}^{\text{media}} \) and it was described in Section 7.4. The value of this timescale is likely to be similar to the generation half-life. Since very little information is available to estimate this value, we chose to sample this parameter in the same way as the parameter \( \tau_{G} \).

In Section 7 we also introduced a parameter \( \nu \) described by equation 7.5. However, as described equation 7.5 was used in a previous version of our simulation model and helped control the stochastic nature of how fast taxpayers that initiate in under-reporting tend towards full evasion. We considered values for this exponent ranging uniformly between 1 and 2 and we found that the trajectories of the amount of income reported by non-compliant taxpayers towards full evasion are not sensitive to the parameter \( \nu \) when it takes integer values larger than 2. Thus, when \( \nu = 2 \) taxpayers that initiated in tax evasion will tend to full evasion according to the half-life time scale \( \tau \). When \( \nu = 1 \) this tendency towards full evasion is dampened and taxpayers that initiate in tax evasion by under-reporting a small fraction of their hideable income take longer to reach the point where they under-report all their hideable income. However, as described in section 7 the current version of the TBM uses a simpler mechanism which qualitatively gives the same results.

\(^{39}\)Let’s denote \( I \) as the income of a typical non-compliant taxpayer and \( R_{n} \) the income s/he reports in his/her tax returns \( n \) years after the audit that resulted in a penalty. Let’s assume that the year after an audit/penalty the taxpayer does not under-report and becomes compliant, thus \( R_{1} = I \). The results reported by Beer et al. [20] imply that \( R_{1} = R_{0} + 2.5 R_{0} \Rightarrow R_{0}/I = 1/3.5 \sim 28.6\% \). The second result implies \( R_{3} = R_{0} + 1.2 R_{0} = (2.2/3.5) I \) then \( R_{3}/I = 0.63 \). This means that 45\% (1/2.2) of what they report after three years (i.e., \( R_{3} \)) is equal to \( R_{0} \) (i.e., what they reported on the year they were found to under-report) and 55\% is the additional amount (i.e., 55\% of \( R_{3} \)) is due to the effect of the audit/penalty.

\(^{40}\)Note that, compared to \( \tau_{G} \) the mode value of \( \tau \) is much closer to the lower bound value than to its upper bound value. This can cause a distortion distribution of \( \tau_{G} \) from its intended form. To compensate and correct for this distortion, we actually extended the range for the upper bound of \( \tau_{G} \).
and does not require this extra parameter $v$.

### 12.4 Survey based Parameters

The parameters listed in table 12.6 were informed by our ALP survey. We way we informed these parameters from the survey is described in the next sections.

<table>
<thead>
<tr>
<th>Name</th>
<th>Symbol</th>
<th>Value</th>
<th>Range</th>
<th>Sampling Distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>bomb.crater.factor</td>
<td>$\kappa_{BC}$</td>
<td>0.58</td>
<td>$[0.50,2.20]$</td>
<td>PERT</td>
</tr>
<tr>
<td>gamblers.fallacy.grad</td>
<td>$\gamma_{GF}$</td>
<td>0.079</td>
<td>$[0.000,0.183]$</td>
<td>PERT</td>
</tr>
<tr>
<td>gamblers.fallacy.intercept</td>
<td>$a_{GF}$</td>
<td>0.159</td>
<td>$[0.000,0.980]$</td>
<td>Derived from $\gamma_{GF}$</td>
</tr>
<tr>
<td>m.qP</td>
<td>$m_x$</td>
<td>0.058</td>
<td>$[0.025,0.116]$</td>
<td>PERT</td>
</tr>
<tr>
<td>s.qP</td>
<td>$s_x$</td>
<td>0.739</td>
<td>$[0.588,0.959]$</td>
<td>Derived from $m_x$</td>
</tr>
<tr>
<td>ave.degree.tTaxes</td>
<td>$\langle k_N \rangle$</td>
<td>2.62</td>
<td>$[2.00,4.00]$</td>
<td>PERT</td>
</tr>
<tr>
<td>beta.personal</td>
<td>$\beta_P$</td>
<td>0.55</td>
<td>$[0.1,0.0]$</td>
<td>PERT</td>
</tr>
<tr>
<td>beta.network</td>
<td>$\beta_N$</td>
<td>0.225</td>
<td>$[0.00,0.5]$</td>
<td>Derived from $\beta_P$ &amp; Inverse</td>
</tr>
<tr>
<td>beta.media</td>
<td>$\beta_M$</td>
<td>0.225</td>
<td>$[0.00,0.5]$</td>
<td>Derived from $\beta_P$ &amp; Inverse</td>
</tr>
<tr>
<td>media.mid.eff ect</td>
<td>$m_m$</td>
<td>0.385</td>
<td>$[0.297,0.472]$</td>
<td>Uniform</td>
</tr>
<tr>
<td>media.steepness</td>
<td>$s_m$</td>
<td>2.245</td>
<td>$[1.792,2.695]$</td>
<td>Derived from $m_m$</td>
</tr>
<tr>
<td>tax.gap.reporting.media.threshold</td>
<td>$\Theta_M$</td>
<td>0.10</td>
<td>$[0.10,0.15]$</td>
<td>Uniform</td>
</tr>
<tr>
<td>c1.dist.weight</td>
<td>NA</td>
<td>0.5</td>
<td>$[0, 1]$</td>
<td>Uniform</td>
</tr>
</tbody>
</table>

Table 12.6: Parameters informed by the ALP survey.
The ALP Survey on Tax Compliance

13 Overview to the ALP Survey on Tax Compliance

The RAND American Life Panel (ALP) is a nationally representative, probability-based panel of over 6,000 U.S. adults, age 18 and older. Panel members are regularly surveyed over the internet for research purposes. As part of their participation, they are provided with the technology to respond to surveys online (a netbook or tablet and high-speed internet access), if they do not already have it. Thus, the panel also includes respondents who would not have had access to the internet otherwise, and who are often excluded from internet-based research. The panel is longitudinal, which allows tracking responses from the same participants over the time and across different surveys administered to them. Since January 2006, ALP has fielded over 450 surveys on topics such as financial decision making, health behavior and outcomes, retirement decisions, inflation expectations, political attitudes related to the presidential elections, working conditions, and more. We used RAND ALP to conduct a survey on tax evasion.

Between July 14th and September 2nd, 2016 we fielded a survey on the ALP assessing perceptions and behaviors regarding federal income taxes. The objectives of this tax evasion survey were to (i) inform our ABM of important behavioral mechanisms that influence the propensity to evade, hence contributing to the structure of the model, and (ii) provide an empirical basis for choosing model-specific parameter values and their uncertainty bounds. The survey results also extended our understanding of how risk and fairness perceptions are influenced by personal experience and vicarious experience through a social network.

Overall, the survey was used to inform the ABM in three ways. First, it provides distributions of agent characteristics, such as whether they are self-employed, as well as their perceived audit and penalty rates. Second, it informs the structure of the social network such as the extent that self-employed taxpayers interact and influence other self-employed taxpayers about taxes. Third, it informs key parameters that drive ABM mechanics. These parameters are listed in table 12.6.

The full survey, including full question wording, is provided in Appendix D, and both survey instrument and data are available on the RAND ALP website at https://alpdata.rand.org/.

This section describes the sample description, overall survey design, and parameters within our ABM on tax evasion that were informed by the survey. This section also follows on from Section 11 and further describes in more detail how the survey was used to inform various the attributes of the agents and the social network structure.

13.1 Sample description

Our target was to obtain data for 1000 respondents. Toward this end, we invited 1320 panel members to participate in our survey. The sample was designed in maximizing overlap with a prior survey on influenza vaccination (labeled Well-being 257), as the Well-being 257 survey contained behavioral questions that bore on free-riding behavior in another domain, which we wanted to compare to similar questions in the tax evasion survey. The sample also oversampled self-employed individuals, since they have greater opportunity to evade taxes than are others. In addition, 50 foreign-born panelists were invited to participate, in order to increase the capacity to gain preliminary insights useful for follow-on research involving cross-national comparisons. Otherwise, the sample was designed to be nationally representative, and raking weights were

41The survey is referred to as Well-Being 456.
provided to account for sampling design and non-response bias. Overall, of the 1320 panelists invited, 1030 completed the survey, for a participation rate of 78.0%.

13.2 Survey design

The introduction of the survey that was presented to the respondents read:

> This survey asks about your experiences and beliefs regarding various taxes and public programs and specifically regarding US federal income taxes. We are only interested in your perspective - there are no right or wrong answers to any of these questions. If you are uncertain about the answer to a question, please give your best estimate. Some of the questions will ask about your perceptions of "people like you." By this, we mean people you think have similar experiences and perspectives, including things like whether and how much they have worked in the past. If you file your tax returns jointly with someone else, please respond according to how your income is reported on those tax returns. As always, your responses will be used only for research purposes, and your individual responses will be confidential.

The survey then contained the following sections.

13.2.1 Current demographics

Demographic variables for all panelists are collected by the ALP quarterly and merged into each survey’s data, so are available for the current analyses. Two of these variables, current job status and current living situation (i.e., married or living with a partner, separated, divorced, widowed, or never married) were critical for this survey, as they affected both question wording and routing among survey questions. As such, for those missing these data in the ALP demographics, we asked these questions again at the beginning of the survey.

13.2.2 Social network characteristics

Respondents were first asked to list the initials of 10 adults that they know, other than their spouse or domestic partner, and interact with on a daily basis. These individuals are referred to as alters. For each alter, respondents were then asked to provide relationship (family, friend, coworker, other), education level, whether they had talked to the alter about taxes in the prior 5 years, and if so how often. In addition, respondents were asked whether each alter was self-employed and whether the respondent knew that the alter had been audited by the IRS in the prior 5 years.

13.2.3 Perceived audit, penalty, income tax and evasion rates

Respondents were asked to use a subjective probability scale (from zero to 100) to estimate the percentage of U.S. taxpayers who are audited by the IRS in a given year. Follow-up questions

---

42 Responses for all probabilistic and proportion questions were collected using a visual slider, anchored at 0 and 100. Similar visual scales have been shown in the past to provide valid probabilistic responses while minimizing response artifacts (e.g., an excess of "50%" responses) found with fill-in probability response modes. As soon as the respondent clicked anywhere on the slider, a dot appeared at that point on the slider and could be moved around. Any slider value was also reflected in a fill-in box below the slider. The fill-in box also allowed the respondent to fill in a response between 0 and 100, which would be reflected on the slider.
asked (a) if an individual only paid [30%, 60%, or 90%, randomly assigned] of their taxes, would their chances of audit be higher, lower, or no different, (b) if higher or lower, what would the new audit probability be, (c) if an individual was audited in the prior year, would their chances of audit this year be higher, lower, or no different, and (d) if higher or lower, would the new audit probability be.

Respondents were then told, "If the IRS detects that a person has underreported their taxes, they will first have to pay the unpaid taxes that were due. In addition, they will be assessed a penalty that is a percentage of the amount they underpaid. This percentage is the penalty rate." They were then asked, "Imagine a person was caught underpaying their taxes by $1000. In addition to having to pay that $1000, how much of a penalty would they have to pay?" Their responses were converted to a percentage value out of the $1000 to obtain the estimate of their perceived penalty rate.

The effective income tax rate was defined for respondents as, "the percent of your income that you owe in taxes to the federal government each year." Respondents were then asked, "What do you think your effective income tax rate was this past year?"

Respondents were finally asked to estimate the proportion of people who intentionally underreported their taxes in a typical year, among (a) all taxpayers in the United States and (b) "100 people like you." A follow-up question posed the hypothetical situation that a widely-disseminated news story came out that half of all US taxpayers underreport on their taxes, asking now how many out of "100 people like you" would underreport. Finally, respondents were asked, "Imagine a person who underreported on their taxes and was audited by the IRS. What is the percent chance that the IRS will catch the underreporting?"

13.2.4 Change in the perceived evasion rate (PER) with the change in tax, audit, and penalty rates

Respondents were randomly assigned to hypothetical situations where the effective tax rate was (a) 50% higher and twice as high or (b) 25% lower or half of what it currently is. Respondents were asked for the evasion rate among "100 people like you" in these hypothetical situations. We refer to these responses as the perceived evasion rate (PER) of the respondent for the different hypothetical situations. We assume that the PER reflects the initiation probability of tax evasion behavior and its magnitude of the respondents even though the survey question was posed in terms of one hundred people that the respondent identifies like themselves rather than self-referring. Respondents were next given hypothetical situations where the audit rate was twice or three times as high as it currently is, and were asked for the evasion rate among "100 people like you" in these hypothetical situations. Finally, they were randomly assigned to hypothetical situations where the penalty rate was (a) 50% higher and twice as high or (b) 25% lower or half of what it currently is. Respondents were asked for the evasion rate among "100 people like you" in these hypothetical situations.

13.2.5 Change in the perceived evasion rate (PER) with the change in tax rate and no deterrence

Respondents were told, "Now finally consider the situation where both audit rate and penalty rate are zero. In this situation, no one is audited and hence no one is penalized. Let's consider how low the effective income tax rate would need to be before everyone reported 100% of their taxes to the IRS, assuming there are no audits or penalties. For this question, assume for the moment that everyone has the same effective tax rate. If each of the effective income tax rates below was
applied to everyone, please indicate if you think (a) the majority of people like you would report their full income OR (b) the majority of people like you would underreport their income." They were then asked to place a check in each row of a table where the rows gradually increased the income tax rate from 1% to 30% and the columns were labeled "Majority of [people/people like you] would report 100% of their taxes" or "Majority of [people/people like you] would under-report their taxes." Half of the respondents received the "people" and half the "people like you" prompts.

13.2.6 Willingness to take a deduction

A hypothetical question asked respondents to consider whether they could claim a $1000 deduction that they were unsure if it was appropriate to take. In addition, they were randomly told that, in the prior year, they had either received a $1000 refund or owed an additional $1000. They were then asked for the percent chance that they would claim this deduction.

13.2.7 Weighting tax fairness-related considerations

Three questions asked respondents to distribute 100 tokens among multiple considerations, according to their relative importance. An automated counter at the bottom of each list helped them keep track. The first list focused on elements of the tax system and included the amount of taxes that they owe, the cost of figuring out their taxes, benefits and public services supported by taxes, and a moral obligation to pay taxes. The second list focused on the relative influence of the social network and media and included their own thoughts on the fairness of the tax system, the thoughts of their family and friends, and what they hear from the media. The third list focused on audits and penalties and similarly included their own thoughts, thoughts of family and friends, and the media.

13.2.8 Perceived value from taxes

A single question asked, "to what extent are the public goods and services that you receive worth the federal income taxes you pay." The repose used a five-point scale and ranged from "Not at all worth it" to "Definitely worth it".

13.2.9 Free-riding

Two sets of five questions assessed each respondent’s perceptions of behaviors that free-ride on a public good. Behaviors included (a) regularly listening to public radio without contributing, (b) illegally copying, downloading, or streaming movies, (c) having a dog but not getting it spayed or neutered, (d) avoiding getting the flu vaccine, and (e) avoiding paying all of the income tax that you owe. The first set of questions assessed personal perceptions and asked if it is "always OK to engage in the described behavior, sometimes OK to do it, or never OK to do that behavior." The second set of questions assessed perceived social norms and asked, "how many people out of 100 would say that it is at least sometimes OK to engage in the described behavior?"

Upon inspecting the survey after fielding, we realized that in multiple rounds of edits, this question mistakenly began with wording that focused on full compliance but had a response mode that focused on majority compliance. We discuss the ramifications of this ambiguity when considering the data for this question.
13.2.10 Influence of public media reporting on overall tax compliance

To estimate the effects of media reporting about overall tax compliance on one’s propensity to intentionally underreport their taxes we asked three questions. The first question asked, "In a typical year, out of all taxpayers in the United States, what percent intentionally underreport their taxes?". We then asked "Now consider people like you. In a typical year, out of 100 people like you, how many intentionally underreport their taxes?". Then, we asked how their answer would change under a hypothetical scenario as follows: "Imagine that a widely-disseminated news story comes out that half of all US taxpayers underreport their taxes. Out of 100 people like you, how many would now underreport their taxes?".

13.2.11 Influence of public media reporting on IRS activities

A single hypothetical question asked, "Imagine that you heard a famous actor was caught and prosecuted for tax evasion. In your mind, would hearing about this make you more or less likely to report all of the taxes you owe to the IRS?". The repose used a five-point scale and ranged from "I would be much more likely to fully report my income" to "I would be much less likely to fully report my income".

13.2.12 Tax and audit experience

Finally, a set of questions asked (a) if ALP data suggested that they were not currently working, whether they had ever worked for pay, (b) if ALP data suggested that they were currently working, whether they worked for someone else or were self-employed, (c) had ever filed a tax return, (d) whether they typically prepare their own tax return or pay someone to do it, (e) whether they had ever been audited, and (f) if married or living with a partner, if that person had been audited in the prior 5 years.

14 Using the ALP Survey to inform the model parameters

In this section, we describe how the analyses of our survey helped inform the simulation model. We start by describing how we informed model parameters. In the subsequent sections, we describe how we used the survey to inform network related properties and how we assigned the initial/baseline behavioral parameters to our taxpayers. The ALP provides population weights for all surveys. However, we have elected to not weigh the current results, given the complexity of some analyses, and instead rely on data from the NDSSL network which is representative for the city of Portland and this was used to inform our taxpayers’ population. Regression models found from our survey analysis were used to inform additional network and behavioral attributes of the taxpayers. These regressions rely on some important covariates such as gender, age and household income that are available in the NDSSL data that is representative for Portland and were used to help assign these additional attributes.

14.1 Assessing the PER with changes in the effective tax rates

Section 7.1 described how taxpayers in our ABM become less compliant as the effective tax rate increases and their perceived deterrence decreases. Figures 7.1 and 7.2 in Section 7.1 provide illustrations of how our taxpayers update their propensity to underreport on their taxes with changing
Figure 14.1: A "slippery slope" representation of how the propensity for tax compliance changes with the effective tax rate and perceived deterrence. Here, this dependence is described by two S-Curves using a CDF of a lognormal distribution for both the dependence on the effective tax rate and perceived deterrence. The value $c_1^{(i)}$, represents the maximum effective tax rate that a taxpayer $i$ remains fully compliant in the hypothetical case of no perceived deterrence. As his/her perceived deterrence increases this threshold maximum tax rate also increases as shown by the value $\tilde{c}_1^{(i)}$. If the effective tax rate is greater than $c_2$, taxpayers will underreport all their hideable income.

14.1.1 Survey questions on the PER with different tax rates

Survey questions 14 and 17 to 20 were used to determine how the PER changes with hypothetical changes in the tax rate. These questions are shown in tables D.4, D.5 and D.6 in Appendix D. First, we used question 14 to obtain the respondent’s PER under the baseline case. A respondent’s PER baseline is the perceived evasion rate that the respondent estimates when the audit rate, penalty rate, and effective tax rate are given by his/her responses to questions 8, 11 and 12 respectively.
shown in table D.4. We then used questions 17 to 20 to determine how the PER changes from the baseline value when the effective tax rate is either increased or decreased with respect to the respondent’s baseline estimate of his/her effective tax rate. Half of the respondents (determined randomly) received questions 17 and 18, which provide hypothetical situations where the effective tax rates are 50% or twice as high as they currently are, and the other half received questions 19 and 20, where the hypothetical effective tax rates are 25% or 50% lower. The introduction to these questions read: "Now let’s consider the effect of changing the effective income tax rate (but the audit and penalty rates remain unchanged). As a reminder, you stated earlier that [PerceivedEvasionRate] out of 100 of people like you underreport their taxes to the IRS."

Because each respondent provided up to three responses within this set of questions, the data were restructured such that each response was treated as a separate observation (i.e., thus, up to three observations per respondent). A baseline observation included the respondents’ baseline estimates of the effective tax rate, penalty rate, audit rate, and evasion rate (among people like themselves). Additional observations corresponded to the estimated evasion rates in the hypothetical scenarios, adjusting the baseline effective tax rate to match the scenario. Table 14.1 provides the three observations for a hypothetical respondent, with their corresponding responses for perceived evasion rate. As you can see, this individual’s baseline perceived effective tax rate was 15%, so for the two observations involving hypothetical increases (of 50% and 100%), the perceived effective tax rate is adjusted upward. The entire restructured data set, with all respondents, is hence designed to be analyzed with the observations (rather than the respondents) as the unit of observation.

<table>
<thead>
<tr>
<th>Question</th>
<th>Perceived Effective Tax Rate (T)</th>
<th>Perceived Penalty Rate</th>
<th>Perceived Audit Rate</th>
<th>Perceived Evasion Rate (E)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>15</td>
<td>40</td>
<td>5</td>
<td>25</td>
</tr>
<tr>
<td>Tax rate 50% higher</td>
<td>17.5</td>
<td>40</td>
<td>5</td>
<td>30</td>
</tr>
<tr>
<td>Tax rate twice as high</td>
<td>30</td>
<td>40</td>
<td>5</td>
<td>50</td>
</tr>
</tbody>
</table>

Table 14.1: Sample respondent data in restructured data set. The rates shown in red were computed based on the question and the baseline response rate. Note that the responses shown have been created for illustration purposes and do not represent actual responses from our survey data.

In total, we started with 1030 respondents and 2918 observations (i.e., not all respondents answered all three questions). Of these, 995 provided baseline estimates for tax, penalty, audit, and evasion rates, resulting in 2902 observations. Two types of response patterns are not used for the current analysis. The first are individuals who gave a zero-perceived evasion rate for all three questions. As we will see below, we presume an S-shaped function and these responses could not be used for the fit. Eliminating these individuals gave us 934 respondents and 2733 observations. The second are those that gave response patterns that were not internally consistent, such as estimating a lower evasion rate in hypothetical questions that presumed a higher effective tax rate. The survey did not enforce consistent responses across questions. These observations were not used for this analysis. Therefore, follow-up responses that were inconsistent with the

---

44 Responses to these questions are, in principle, relative to a respondent’s initial perceived evasion rate. Hence, respondents who gave an initial evasion rate estimate of 0% were automatically routed into the first two questions, and respondents who gave an initial evasion rate estimate of 100% were automatically routed into the second two questions.
baseline perceived evasion rate are excluded. Considering only the consistent responses for the 934 respondents gives us 2483 observations in the final analytic dataset or 85.1% of the original data. As before, because responses were provided using a clickable number line (between 0 and 100), imprecision in clicking could cause an error in response. Thus, to retain as many responses as possible, we allowed responses that were inconsistent by up to 5%, recoding these as no change (i.e., the same as the prior response).

### 14.1.2 Data exploration

Figure 14.2 uses all available data with 2918 observations to show how perceived evasion rate varies with perceived effective tax rate with lighter contours representing greater density of responses. The red line indicates the best fit OLS regression estimate, constrained to have a zero intercept. As can be seen, perceived evasion rates generally increase with perceived effective tax rates, as is expected. Note that this analysis treats all observations as independent observations. A better indication of the dependence of PER on the perceived effective tax accounts for the dependence of the observations on the respondents. This dependence can be found using a panel level linear regression approach [45]. We applied both a fixed and a random effects panel regression model to the filtered dataset contains 2483 observations to find the elasticity of the PER on the perceived effective tax rate. This elasticity was found to be 0.997 for the fixed-effects and 0.964 with the random-effects models. This shows that within the set of responses for each respondent, the PER depends linearly on the perceived effective tax rate.

### 14.1.3 Fitting an S-Curve for the dependence of the PER with tax rate

The analysis on elasticity assumes that dependence is described as $E = \kappa T^{\epsilon_T}$, where $E$ denotes the PER, $T$ denotes the perceived effective tax rate, $\kappa$ is a constant and $\epsilon_T$ denotes the constant elas-
ticity. However, the assumption of a constant elasticity over the full range of perceived effective tax rates is unlikely to be correct. Instead, to inform our ABM we assume that the dependence of $E$ on $T$ over the full range of perceived tax rates is best described by an S-Curve. An S-Curve is an attractive description for several reasons. First, logically the perceived evasion rate should be zero percent for a zero tax rate (and presumably some range above that), and above that it should increase smoothly to some point above which the evasion rate will be 100%.

Second, an S-Curve can be easily estimated with few parameters using standard statistical distributions. Figure 14.3 illustrates the process of fitting a respondent’s PERs for different tax rates using an S-Curve function. In this figure, the S-Curve is described by two parameters $m$ and $g_m$ instead of $m$ and $s$. The parameter $g_m$ is related to both $m$ and $s$ and represents the gradient of the S-Curve when the effective tax rate is equal to $m$.

Footnote 45: Here we presume that the evasion rate will be 100% at any point above the theoretical point in the Laffer curve where there is no longer an incentive to work for pay, estimated to be approximately 70%.
Our approach estimated an S-Curve model for each respondent, using their three responses and a hypothetical fourth defined by the Laffer curve which in Figure 14.3 we refer to as a fictitious point as it is not informed by the respondent. Together, these four points are shown in dark blue in the figure. This provides, for each respondent, two parameters ($s$ and $m$) that define his/her S-Curve. We then determined the overall distribution of $s$ and $m$ across respondents, describing it with the median and inter-quartile range (IQR). Finally, we took the tangent to the S-Curve at the median value of the distribution of $m$ to obtain the value for $g_m$.

We considered two functional forms for the S-Curve. The first was the cumulative distribution function (CDF) of a log-normal distribution and the second a logistic function. However, for the purpose of informing the ABM, we focused on the former. In order to run our analyses to estimate the best S-Curve fit, we had to transform the data to a linear form. The functional forms of the S-Curves and the corresponding data transformations to a linear form are described in Appendix E.

Our analysis can be described as a manual implementation of a fixed effects panel linear regression model. To support the results obtained by this analysis we also ran both a standard fixed effects and a random effects panel linear regression model using both STATA and the \texttt{plm} package in R. We verified that the S-Curve fits produced by these regressions gave similar results to those presented here.

<table>
<thead>
<tr>
<th></th>
<th>Min.</th>
<th>1st Qu.</th>
<th>Median</th>
<th>3rd Qu.</th>
<th>Max.</th>
<th>Mean</th>
<th>sd</th>
</tr>
</thead>
<tbody>
<tr>
<td>$s$</td>
<td>0.13</td>
<td>1.80</td>
<td>2.37</td>
<td>3.18</td>
<td>11.58</td>
<td>2.71</td>
<td>1.45</td>
</tr>
<tr>
<td>$m$</td>
<td>0.00</td>
<td>0.25</td>
<td>0.34</td>
<td>0.44</td>
<td>0.86</td>
<td>0.35</td>
<td>0.15</td>
</tr>
</tbody>
</table>

Table 14.2: Summary statistics for distribution access the respondent of parameters $s$ and $m$ describing log-normal S-Curve fits for the dependence of the PER with increasing tax rates.

Table 14.2 displays the summary statistics of the distribution of values for the parameters $s$, and $m$ that described the log-normal S-Curve fits. Using median and interquartile values from the table, we show in Figure 14.4 our range of log-normal S-Curve fits. The S-Curve in red uses the median values of the parameters. The dotted blue curves show the S-Curves using the 1st and the 3rd quantile values of these parameters. Thus, the blue dotted curves illustrate the large uncertainty range in the fit. Despite this large range, the gradient of the tangent lines to the S-Curves at the point where the PER is 50% that shown in green in the plot is very similar. This is because the values of $s$, $m$ and $g_m$ are related and $g_m$ is given by

$$g_m = \frac{s}{m\sqrt{2\pi}}$$

(14.1)

and results in a low range of the gradient value $g_m$ that are very close to the median value of 2.9. Therefore, we can assume that our respondents can each be described by different S-Curves with a different $m$ value but with the same gradient parameter $g_m = 2.9$. Projecting the tangent lines to the S-Curves down to zero PER allowed us to find an estimate of the value for $c_i^{(i)}$ of the respondent. This represents the approximate threshold tax rate when the perceived evasion rate (in people like you) begins to increase from a 0% level under typical (i.e., baseline) perceived levels of deterrence. We will refer to this method as a “parallel projection”. Mathematically, this method

\footnote{The reason we used a manual implementation of a fixed effects panel linear regression model is that we could more readily extractive distribution of the parameters $s$ and $m$ for each respondent. Instead, STATA and R produced overall results of the regression model and extracting the information of the distribution of fits across the respondents is less easily obtained. Please refer to Appendix F for a summary of the fixed effects and a random effects panel linear regression model run in STATA.}
Figure 14.4: Lognormal S-Curve fits using the parameters and inter-quartile range (IQR) given in Table 14.2.

is described by using the equation for the line \( E = g_m T + k \) separately for each respondent. Then knowing that \( E = 0.5 \) (i.e., the PER is 50%) when the effective tax rate \( T = m \) allows use to find the intercept value, which is given by \( k = 0.5 - g_m m \). Thus, using the different intercept values for each respondent we estimated their \( \tilde{c}_1(i) \) by

\[
\tilde{c}_1(i) = \frac{k}{g_m} = \frac{2g_mm^{(i)} - 1}{2g_m}.
\]

This method provided us with an estimate for \( \tilde{c}_1(i) \) for the respondent’s baseline perceived deterrence rate. As shown in Figure 14.3 by the light blue and light red points, each respondent provided estimates for the PER for hypothetical scenarios where the perceived deterrence rate was different from the baseline value. Using this same method of parallel projection we estimated the values of the respondent’s \( \tilde{c}_1(i) \) values for these different perceived deterrence rates. Thus we found how the PER depends on the perceived deterrence rate which we have denoted by \( \tilde{x}^{(i)} \). Mathematically,

\[
\tilde{c}_1(i) (\tilde{x}^{(i)}) = T^{(i)} - \frac{E_{\tilde{x}^{(i)}}}{g_m},
\]

where \( T^{(i)} \) is the respondent’s perceived effective tax rate and \( E_{\tilde{x}^{(i)}} \) is the respondents PER under the hypothetical scenario where his/her perceived deterrence rate is \( \tilde{x}^{(i)} \). This analysis is explained in more detail in the next sections.
14.2 Assessing the PER with changes in the perceived deterrence rate

14.2.1 Survey questions on the PER with different perceived deterrence rates

Questions 21 to 26 of our survey were designed to determine how the PER changes with hypothetical changes in the audit and penalty rates. All respondents received the two questions on audit rate (21 and 22).

Table 14.3 below extends the logic of Table 14.1 for our hypothetical respondent, now including not only data from those questions directly related to how compliance shifts with effective tax rate, but also how compliance shifts with audit and penalty rates. As before, the penalty and audit rates are adjusted to reflect the question wording, and the perceived evasion rates are the responses to those questions. Overall, respondents gave 6460 responses, out of a possible 7210 (94.5%).

<table>
<thead>
<tr>
<th>Question</th>
<th>Perceived Effective Tax Rate (T)</th>
<th>Perceived Penalty Rate</th>
<th>Perceived Audit Rate</th>
<th>Perceived Evasion Rate (E)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>15</td>
<td>40</td>
<td>5</td>
<td>25</td>
</tr>
<tr>
<td>Tax rate 50% higher</td>
<td>17.5</td>
<td>40</td>
<td>5</td>
<td>30</td>
</tr>
<tr>
<td>Tax rate twice as high</td>
<td>30</td>
<td>40</td>
<td>5</td>
<td>50</td>
</tr>
<tr>
<td>Penalty rate 50% higher</td>
<td>15</td>
<td>60</td>
<td>5</td>
<td>15</td>
</tr>
<tr>
<td>Penalty rate twice as high</td>
<td>15</td>
<td>80</td>
<td>5</td>
<td>10</td>
</tr>
<tr>
<td>Audit rate twice as high</td>
<td>15</td>
<td>40</td>
<td>10</td>
<td>20</td>
</tr>
<tr>
<td>Audit rate three times as high</td>
<td>15</td>
<td>40</td>
<td>15</td>
<td>15</td>
</tr>
</tbody>
</table>

Table 14.3: Sample respondent data in a restructured dataset. The rates shown in red were computed based on the question and the baseline response rate. Note that the responses shown have been created for illustration purposes and do not represent actual responses from our survey data.

However, we filtered the data by running a few consistency checks. Specifically, we imputed values for logical skips, dropped respondent if they had missing baseline data or if baseline audit rate was larger than 30%. Our final analytic dataset included 822 (79.8%) respondents with 4933 responses (68.4%).

In addition to questions 21 to 26, the survey included a further question (question number 27), that was designed to determine the PER of the respondents under the hypothetical scenario of no deterrence (i.e., assuming there are no audits or penalties). However, we note that this question had a problem as it starts off by asking the respondents to consider how low the effective income tax rate that would need to be before everyone reported 100% of their taxes to the IRS, assuming there are no audits or penalties, but then as for the rate where the majority of people would be compliant. In our analysis, we interpret the responses as being the tax rate where the respondents think that the PER increases above the 50% mark under the hypothetical scenario of no deterrence.

\[\text{Equation for the distribution for the tax rate at which underreporting would begin under zero deterrence was highly concentrated near zero, falling monotonically, with most of the weight below 10\%. Because question 27 was ambiguous in terms of whether respondents interpreted their response as reflecting the point at which all or 50\% of the population would underreport, we analyzed the data under both assumptions. The results under the 50\% assumption closely mirrored our a priori theoretical expectations, as was our initial intention, whereas the 100\% assumption pro-}

\[\text{Our expectation for the distribution for the tax rate at which underreporting would begin under zero deterrence was highly concentrated near zero, falling monotonically, with most of the weight below 10\%. Because question 27 was ambiguous in terms of whether respondents interpreted their response as reflecting the point at which all or 50\% of the population would underreport, we analyzed the data under both assumptions. The results under the 50\% assumption closely mirrored our a priori theoretical expectations, as was our initial intention, whereas the 100\% assumption pro-}

76
14.2.2 Data exploration

In figure 14.2 we showed a contour plot of how PER changes for different tax rates using the full unstructured data set that presumes that all observations are independent. A similar analysis on how the PER changes for different hypothetical cases with different deterrence rates revealed no dependence of the PER on the perceived deterrence rates. However, after arranging the data in a panel format as presented in table 14.3, we analyzed the data to estimate the elasticities. The functional form of the dependence of the PER with tax rate and perceived deterrence is

\[ E = \kappa T^{\epsilon_T} q^{\epsilon_q} P^{\epsilon_P}, \] (14.4)

whereas before \( \kappa \) is a constant and \( \epsilon_T, \epsilon_q \) and \( \epsilon_P \) denotes the constant elasticities with respect to the tax rate, audit rate, and penalty rate. We transformed the data by taking the logarithms of both sides of equation 14.4 and carried out a panel level linear regressions. Using a fixed-effects model we found that the elasticities were 0.997, -0.63 and -0.62 respectively, and the dependence was found to be statistically significant. This analysis is further described in Appendix F. This simple analysis revealed a clear dependence of the PER on the perceived audit rate and penalty rate. Furthermore, the values for the audit and penalty rate elasticities were similar. This helped support our approach that considers the product of the perceived audit and penalty rates as our measure of the perceived deterrence rate. Table 14.4 shows the quartile information of the perceived deterrence rate of our respondent.

<table>
<thead>
<tr>
<th>Min.</th>
<th>1st Qu.</th>
<th>Median</th>
<th>3rd Qu.</th>
<th>Max.</th>
<th>Mean</th>
<th>sd</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.000</td>
<td>0.007</td>
<td>0.020</td>
<td>0.060</td>
<td>7.500</td>
<td>0.080</td>
<td>0.281</td>
</tr>
</tbody>
</table>

Table 14.4: Quartiles of the perceived deterrence rate.

14.2.3 Extracting the \( \tilde{c}^{(i)}_1 \) values for the different perceived deterrence rates

Following on from the approach described in Section 14.1.3, we used our value for the gradient \( g_m \) and the responses to questions 21 to 26 to estimate how \( \tilde{c}^{(i)}_1 \) depends on the perceived deterrence rate \( \tilde{x}^{(i)} \). The method we used is illustrated in figure 14.3. For each respondent we used equation 14.3 and the constant gradient \( g_m \) to find the parallel line that goes through the respondents value for \( E \) and \( T \) for the different cases. This allows us to estimate the respondent’s \( \tilde{c}^{(i)}_1 \) for different perceived deterrence rates \( \tilde{x}^{(i)} \). Therefore, for each respondent we obtained a set of five \( \tilde{c}^{(i)}_1 \) values.

14.2.4 Extracting the \( c^{(i)}_1 \) value for a zero perceived deterrence

Table 14.5 shows the how our 822 respondents answered question 27. Respondents who provided a tax range response between 30 and 100% had their value modified to a new value that ranged between 30 and 70%. This modification satisfied a criterion that aimed to create a smooth distribution of the responses. Moreover, we then ran a second set of modifications which ran consistency

\[ \text{distribution that was unrealistically high. This analysis increased our confidence that most respondents used the 50% interpretation, and our analyses reported below make this assumption.} \]

\[ ^{48}\text{The first } \tilde{c}^{(i)}_1 \text{ is obtained by using the data presented by the first three rows of table 14.3. These also determine the gradient value } g_m. \text{ The other four rows provide the } \tilde{c}^{(i)}_1 \text{ values for the different hypothetical cases with different deterrence rates.} \]
Table 14.5: Data showing the number of respondents thinking that the critical compliance decisions would occur between the different ranges of tax rates.

<table>
<thead>
<tr>
<th>Tax Rate Range</th>
<th>Counts</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-1 %</td>
<td>35</td>
<td>4.4%</td>
</tr>
<tr>
<td>1-2.5 %</td>
<td>8</td>
<td>1%</td>
</tr>
<tr>
<td>2.5-5 %</td>
<td>19</td>
<td>2.4%</td>
</tr>
<tr>
<td>5-10 %</td>
<td>81</td>
<td>10.2%</td>
</tr>
<tr>
<td>10-15 %</td>
<td>160</td>
<td>20.4%</td>
</tr>
<tr>
<td>15-20 %</td>
<td>161</td>
<td>20.2%</td>
</tr>
<tr>
<td>20-25 %</td>
<td>110</td>
<td>13.9%</td>
</tr>
<tr>
<td>25-30 %</td>
<td>35</td>
<td>4.4%</td>
</tr>
<tr>
<td>&gt;30%</td>
<td>182</td>
<td>23%</td>
</tr>
</tbody>
</table>

Figure 14.5: Histogram and smoothed distribution (red) of the modified responses to question 27. The histogram uses a color code shown in the legend which is used to help illustrate how the respondents actually answered question 27 (also shown in table 14.5). Responses to question 27 were modified based on the consistency checks and responses to questions 17 to 26 as described in the main text. The histogram shows the results after this modification and is used to provide the effective tax rate where the PER crosses the 50% value. The distribution shown in red provides an approximate smoothed fit to the histogram.

These consistency checks can be explained by referring back to figure 14.3. It is possible that the value for $c_{1}^{(i)}$, found using the method of parallel transportation by running a line through the green point found from the response to question 27, is greater than (i.e., to the right of) any one of the $c_{1}^{(i)}$ found using the parallel transportation by running a line through the points that reflect the responses to questions 17 to 26. However, by definition $c_{1}^{(i)}$, that occurs under zero perceived deterrence must be smaller than any of the $c_{1}^{(i)}$ that occur when the perceived deterrence is non-zero. Thus, when this happens $c_{1}^{(i)}$ is decreased and is substituted by the minimum value of the set of $c_{1}^{(i)}$ found using questions 17 to 26 and parallel transportation.

An example may further help explain these consistency checks and modifications. Consider the case where a respondent’s baseline PER is greater than 50% and his/her effective tax rate is smaller than 30%. Further, consider that this respondent answered question 27 and thinks that with no perceived deterrence the effective tax rate must be
smoothed distribution of the responses to question 27 after applying our consistency-based corrections. Under the hypothetical case of zero perceived deterrence, our analysis is shown in Figure 14.6.

Figure 14.6: Sampling distribution for $c_1^{(i)}$. The probability distribution shown in green was found from our analysis of the ALP data for question 27. The probability distribution shown in blue is taken from a beta distribution that monotonically falls and reaches zero when $c_1^{(i)}$ is 10%. This distribution was suggested based on our expert consultation. Our survey-based distribution also falls monotonically but has a longer tail.

Section 14.5 provides an estimate of the distribution of the effective tax rate where the PER increases above the 50% mark. However, in our ABM $c_1^{(i)}$ is defined as taxpayer’s $i$ maximum effective tax rate that sustains full compliance (i.e., where his/her PER is zero or negligible) under zero perceived deterrence. To estimate $c_1^{(i)}$ we apply the parallel projection method as illustrated in Figure 14.3. For about one-third of our respondents, the parallel projection method resulted in a negative value $c_1^{(i)}$ estimate. For these respondents, we assumed their $c_1^{(i)}$ was zero. Figure 14.6 shows the probability distribution for sampling the values of $c_1^{(i)}$ using our approach. It also shows an alternative distribution for $c_1^{(i)}$ that was suggested based on our expert consultation. In our ABM we select $c_1^{(i)}$ by sampling from these two distributions. The model parameter $c1.dist.weight$ referred to in Table 12.6 provides the weights we place on using the ALP based distribution.

### 14.3 Model for Perceived deterrence: Estimating the parameters $m_x$ and $s_x$

Using the parallel projection method that was illustrated in Figure 14.3, we obtained five different $\tilde{c}_1^{(i)}$ points for different deterrence levels, and one $c_1^{(i)}$ point for each respondent. In addition to these six points, we set a "fictitious" seventh point at $\tilde{c}_1^{(i)} = c_2$ for a perceived deterrence rate $\tilde{x}^{(i)}$ equal to 100% $^{50}$ As illustrated in Figure 14.7 these seven points form a trajectory of $c_1^{(i)}$ values with increasing perceived deterrence rate $\tilde{x}^{(i)}$. The quantity $\Phi(\tilde{x}^{(i)})$ was defined in Section 7.1.

---

$^{50}$This is an arbitrary large value that we chose. A perceived deterrence rate of 100% is achieved for example when the perceived audit and penalty rates are respectively 10% and 1000%. This is certainly possible, but our data showed that very few respondents had perceived deterrence rates close or above 100%.
Figure 14.7: An illustration of a respondent’s $c_1^{(i)}$ and $\tilde{x}^{(i)}$ points for different perceived deterrence rates. The y-axis on the left hand side shows $\Phi(\tilde{x}, m_x, s_x)$ which represents the proportion of the distance between $c_1^{(i)}$ and $c_2$ and ranges between 0 and 1. The y-axis on the right hand side shows $\tilde{c}_1^{(i)}$, which varies between $c_1^{(i)}$ and $c_2$. Even if each individual has a different starting value of $c_1^{(i)}$, the model assumes each taxpayer is described by the same function $\Phi(\tilde{x}, m_x, s_x)$. To fit the S-Curve we assumed that each respondent had a seventh fictitious point (not shown) at $\tilde{x}^{(i)} = 100\%$. These seven points define a trajectory. The red S-Curve show the best fit to our data on for $\Phi(\tilde{x}, m_x, s_x)$ used by the model. The dotted blue lines show the range of S-Curve fits described by the IQR is shown in Table 14.6.

by equation 7.2 and it allows us to represent the trajectory of the $c_1^{(i)}$ values as a proportion of the interval between the respondent’s $c_1^{(i)}$ value to $c_2 = 70\%$. We used these trajectories to fit a log-normal S-Curve model to each respondent’s data using the same approach we used in our analysis of how the PER depends on the effective tax rate. For each respondent, this gave us two parameters $(s_x$ and $m_x)$ that define each respondent’s S-Curve. We then determined the overall distribution of $s_x$ and $m_x$, across respondents from which we found the median and inter-quartile range (IQR). As described previously, this approach is basically a manual implementation of a fixed-effects panel linear regression model. As before, to support and verify the results obtained by this analysis we also ran both fixed- and random-effects panel linear regression models using STATA and the plm package in R to estimate the best S-Curve fit to the data. . Table 14.6 displays

<table>
<thead>
<tr>
<th></th>
<th>Min.</th>
<th>1st Qu.</th>
<th>Median</th>
<th>3rd Qu.</th>
<th>Max.</th>
<th>Mean</th>
<th>sd</th>
</tr>
</thead>
<tbody>
<tr>
<td>$s_x$</td>
<td>-8.176</td>
<td>0.588</td>
<td>0.739</td>
<td>0.959</td>
<td>5.378</td>
<td>0.823</td>
<td>0.711</td>
</tr>
<tr>
<td>$m_x$</td>
<td>0.000</td>
<td>0.025</td>
<td>0.058</td>
<td>0.116</td>
<td>2.584</td>
<td>0.115</td>
<td>0.213</td>
</tr>
</tbody>
</table>

Table 14.6: Summary statistics for distribution of parameters $s_x$ and $m_x$ describing log-normal S-Curve fits for the dependence of the $c_1^{(i)}$ on perceived deterrence rate $\tilde{x}^{(i)}$. 

80
Questions 13, 14 and 15 were used to determine how taxpayers may react and change their PER when news provided by the media reports the aggregate level of evasion in the population. As a reminder, question 14 was used to get the baseline estimate of the PER for each respondent (i.e., the perceived evasion rate out of 100 people like you). Question 13 asked the respondents to estimate the percent of taxpayers that intentionally underreport their taxes in a typical year. Then, Question 15 asked the respondents to consider the scenario where a widely-disseminated news story came out reporting that half of all US taxpayers underreport their taxes. Respondents were then asked to estimate the percent of taxpayers, similar to themselves that would intentionally underreport their taxes under this new hypothetical scenario. In our analysis of the responses to these questions, we make a key assumption that the relationship between PER in the population

\[ \text{Change in PER in People like you} = \text{Change in PER in the Population} \times \text{(some function of the parameters)} \]

Table 14.7: Summary statistics for distribution of the parameters \( s_m \) and \( m_m \) describing log-normal S-Curve fit.

<table>
<thead>
<tr>
<th></th>
<th>Min.</th>
<th>1st Qu.</th>
<th>Median</th>
<th>3rd Qu.</th>
<th>Max.</th>
<th>Mean</th>
<th>sd</th>
</tr>
</thead>
<tbody>
<tr>
<td>( m_m )</td>
<td>0.00</td>
<td>0.40</td>
<td>0.47</td>
<td>0.55</td>
<td>1.12</td>
<td>0.49</td>
<td>0.14</td>
</tr>
<tr>
<td>( s_m )</td>
<td>-104.56</td>
<td>2.05</td>
<td>2.69</td>
<td>3.66</td>
<td>125.32</td>
<td>11.48</td>
<td>27.93</td>
</tr>
</tbody>
</table>

51See Appendix D for the questions.
Figure 14.9: Illustration of how respondent’s perceived evasion rate (PER) in people like you compares to their perceived evasion rate in the population. The points illustrate how some respondents answered questions 13 to 15. Each respondent has two points, a baseline point as informed by question 13 and 14, and a hypothetical point as informed by question 15. Points that fall below the dotted diagonal line and in the shaded area represents cases where the respondent believes to be more compliant than most people in the population. Points that fall above the dotted diagonal line represents cases where the respondent believes to be less compliant than most people in the population. The red dotted line provides the best log-normal S-Curve fit to the whole data set (not just to the ten points shown).

and the PER in people like you is monotonically increasing and follows an S-Curve. Therefore, an increase/decrease in the PER in the population would be reflected in an increase/decrease in the PER in people like you. However, because a subset of our respondents (9.8%) violated this monotonicity assumption, their data was not used to estimate the assumed S-Curve. The number of respondents who answered questions 13 to 15 was 1001. Out of these, 903 passed our consistency check for monotonicity and were used for the S-Curve fit.

Figure 14.8 illustrates the how respondents answered these questions and which ones violated this monotonicity assumption. A minority of cases (9.8%) shown by the dimmed gray points shown in the off-diagonal quadrants failed the consistency check and violated the monotonicity assumption. These were removed from our analysis for fitting an S-Curve. However, given the clickable response mode to this question, it is possible that respondents made small errors and did not intend to violate the monotonicity assumption. Therefore, we allowed for some error. This can be seen in the illustration by the inclusion of some off-diagonal points shown in bold red if they were within 5% of fulfilling our strict check for monotonicity. These points that marginally violate the monotonicity assumption were included and were left unmoved in the analysis for fitting our S-Curve.

Figure 14.9 provides an illustration of how the data that passed the consistency check for monotonicity was analyzed to fit the S-Curve. On the x-axis, we show the respondents perceived evasion rate in the population. On the y-axis, we show the PER in people like you. For each respondent, we have two points with each pair designated by its own color. For illustration, con-
Figure 14.10: Illustration of the combined effects on compliance with changes in tax rates and the PER in population. Here we consider a perceived deterrence rate of 2.0% as informed by table 14.4. The importance of the PER in population on compliance is 40% of the importance of changes in the effective tax rate. This value is equal to the ratio between the mode values of $\beta_M$ and $\beta_P$ that are respectively 22.5% and 55%.

Consider the two circled pink points. The first point (labeled Actual) uses the responses to question 13 and 14 as our x and y coordinates respectively. The second point (labeled Hypothetical) uses the response to question 15 as a y coordinate and uses 50% for all respondents as our x-coordinate, marking the point where half of all US taxpayers underreport their taxes. Using the panel-level regression approach of fitting an S-Curve to data as described previously, we found the parameter values that best fit the log-normal S-Curve. Table 14.7 shows the summary statistics for the distribution of parameters $s_m$ and $m_m$ describing our log-normal S-Curve fit.

One concern with the survey-derived estimates is that, whereas they are direct estimates of psychological reactions, they are based on self-report data, which could be subject to various biases in reporting. To provide more robust estimates, we conducted a similar analysis with very different data, those obtained from Google Trends. We used Google Trends to obtain search activities on the internet using terms associated with tax evasion change for different countries with very different tax compliance rates. These data have the advantage of being more objective, but are themselves limited by search activity is only a proxy for the effect of media on psychological perceptions and having to do this analysis across multiple countries. We carried out a similar S-Curve fit analysis of the data to extract the parameters $m_m$ and $s_m$. The estimated values for $m_m$ and $s_m$ in that analysis were 0.297 and 1.792 respectively. Our analysis of data obtained by Google Trends searches has been documented in Appendix 14.7 in a more complete manner.

We decided to uniformly sample the parameters $m_m$ and $s_m$ between the values given by our Google Trends analysis and the median values shown in table 14.7. Both these analyses and S-Curve plots show that the effect of news reports about compliance levels at the population level is negligible when the PER (or the tax gap) in the population is smaller than a threshold value.
that ranges between 10% and 15%. Figure [14.10] provides a 3D plot of how the S-Curves for compliance with changing tax rates and for increasing level of PER in the population. Unlike the 3D plot in figure [14.1] showing how compliance changes with tax rate and perceived deterrence, here the PER in the population is correlated to compliance. Therefore in figure [14.10] we note that when the tax rate is very high the PER in the population would also very likely be high and thus the surface at the near corner does not occur in the simulation model. The weights used to describe the contribution of the effective tax rate and PER in the population on compliance are obtained from the ALP data and the analysis described in the next section.

14.5 The Weights: model parameters $\beta_P$, $\beta_N$ and $\beta_M$

Questions 30 and 31 of our survey separately asked respondents to distribute 100 tokens among personal, social and media-related considerations when respectively considering the fairness of the taxes they pay and the risks involved in being audited and penalized. Figure [14.11] shows a box-plot showing the distribution of these considerations amongst the respondents. These distributions helped inform the values for the weights $\beta_P$, $\beta_N$ and $\beta_M$ used in our model as reported in table [12.6]. Although our model does distinguish between the fairness and the risk considerations, figure [14.11] suggests that these distributions were similar. Therefore, we chose to consider one unique set of weight parameters that apply to both types of considerations. From this analysis, we concluded that respondents hardly made any distinctions between social interactions and media feedbacks. Therefore, for our model, we chose that the values for $\beta_N$ and $\beta_M$ are always equal to each other. We note that a separate analysis by Alm et al. in 2009 [9] that used a laboratory experiment approach, also found that the network and media considerations carry equal weight but are lower than what is suggested by our analysis of the survey data and are closer to 11% each [52]. Since the sum of the weights adds to one, the sampled value of $\beta_P$ determines the values for $\beta_N$ and $\beta_M$. Therefore, the important parameter to consider is $\beta_P$ and distribution of tokens for personal considerations. Figure [14.12] shows the actual distributions for the personal considerations only.

In our calibration and sensitivity analysis of our ABM, we consider multiple model realization cases and for each, we use different values for the weights $\beta_P$, $\beta_N$ and $\beta_M$. In any one single realization case of our ABM, these weights do not vary across taxpayers but rather take specific values that apply to all taxpayers in the model and represent the population average weight of these different considerations. However, according to our sampling method and experimental design, these weights change across model realization cases. This allows us to consider different scenarios whereby the population places different weights on these considerations. An alternative approach considers a population where each taxpayer was assigned a different attribute value for the weights $\beta_P$, $\beta_N$ and $\beta_M$ and where $\beta_P$ is taken from a combination of the first two distribution shown in Figure [14.12]. However, this alternative would have fixed the mean population values of these effects for all possible model realization cases. Thus, it would not have easily allowed us to explore the effects on the model outcomes of changing the weights placed on these considerations. To sample the $\beta_P$ parameter value that is representative of the population average weight for a given model realization we used a PERT distribution ranging from 0 to 100% and with a mean of 55% taken from the analysis of questions 30 and 31. The third histogram in Figure [14.12] shown in gray illustrates the PERT distribution used.

[52] Professor Alm kindly shared the data he collected in his 2009 paper with us. Appendix B describes some related analyses we carried out using his data.
Figure 14.11: Box-plot showing the minimums, maximums, and IQRs of the distribution of tokens respondents gave to the different levels considerations and for questions 30 and 31 that respectively focused on tax fairness and risks of being audited and penalized. Means are shown as black points while medians as horizontal black lines.

Figure 14.12: Histograms showing the distribution of the tokens assigned to personal-level considerations by the respondents in questions 30 and 31 which respectively relate to fairness (blue) and risk (orange). The third histogram in on the right is shown in gray provides shows the PERT distribution of weights sampled for our model runs.
14.6 The Gamblers Fallacy Effect

Question 9a and 9b of our survey were used to describe the *Gambler’s Fallacy Effect*. Respondents were randomly split into three groups and each group respectively had to consider the audit probability for taxpayers that reported 30, 60 or 90% of their income. We first asked the respondents whether they thought that the risk of being audited was greater, lower or stayed the same compared to the baseline value, which they had provided by answering question 8. We then asked them to estimate the new audit probability. Tables 14.8 and 14.9 summarize the results we found by providing summary statistics of the new audit probability expressed as a proportion of their baseline value. We refer to this proportion as the Gamblers Fallacy multiplicative magnitude. Table 14.8 provides a first summary of the results by showing on each row the number of respondents that thought that the audit probability would be higher, the same or lower with respect to the baseline value. Table 14.9 summaries the same results but here each row represents one of the different groups. Very few respondents believed that the risk of being audited would decrease. Most respondents thought that the risk would either increase or stays the same. The multiplicative magnitude of the increase weakly depends on the proportion of income reported. To inform

<table>
<thead>
<tr>
<th>Change</th>
<th>Min.</th>
<th>1st Qu.</th>
<th>Median</th>
<th>3rd Qu.</th>
<th>Max.</th>
<th>Mean</th>
<th>NA’s</th>
<th>N</th>
<th>N.prop</th>
</tr>
</thead>
<tbody>
<tr>
<td>higher</td>
<td>0.00</td>
<td>1.50</td>
<td>2.00</td>
<td>3.75</td>
<td>313</td>
<td>138</td>
<td>1</td>
<td>498</td>
<td>0.49</td>
</tr>
<tr>
<td>same</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1</td>
<td>1.00</td>
<td>0</td>
<td>454</td>
<td>0.45</td>
</tr>
<tr>
<td>lower</td>
<td>0.00</td>
<td>0.33</td>
<td>0.62</td>
<td>0.80</td>
<td>1</td>
<td>0.59</td>
<td>0</td>
<td>61</td>
<td>0.06</td>
</tr>
</tbody>
</table>

Table 14.8: Summary of question 9b for all records where each row represents the way the respondent answered question 9a. The values in the columns ranging from the Min. to the Mean represent the new audit probability expressed as a proportion of the respondent’s baseline audit probability.

<table>
<thead>
<tr>
<th>Reported Income (%)</th>
<th>Min.</th>
<th>1st Qu.</th>
<th>Median</th>
<th>3rd Qu.</th>
<th>Max.</th>
<th>Mean</th>
<th>NA’s</th>
<th>N</th>
<th>N.prop</th>
</tr>
</thead>
<tbody>
<tr>
<td>30%</td>
<td>0.00</td>
<td>1</td>
<td>1.21</td>
<td>2.50</td>
<td>313</td>
<td>4.17</td>
<td>5</td>
<td>349</td>
<td>0.34</td>
</tr>
<tr>
<td>60%</td>
<td>0.00</td>
<td>1</td>
<td>1.00</td>
<td>2.25</td>
<td>31300</td>
<td>95.96</td>
<td>9</td>
<td>349</td>
<td>0.34</td>
</tr>
<tr>
<td>90%</td>
<td>0.00</td>
<td>1</td>
<td>1.00</td>
<td>1.75</td>
<td>31300</td>
<td>106.70</td>
<td>4</td>
<td>332</td>
<td>0.32</td>
</tr>
</tbody>
</table>

Table 14.9: Summary 9b where the rows represent the three groups of respondents that were asked to respectively consider the audit probability for taxpayers that reported 30, 60 or 90% of their income. The values in the columns ranging from the Min. to the Mean represent the new audit probability expressed as a proportion of the respondent’s baseline audit probability.

the the parameters $\gamma_{GF}$ and $\alpha_{GF}$ we re-express equation (7.16) as

$$
\log(1 + \kappa_{GF}^{(i)}) = \gamma_{GF} \log(1 - a_{t}^{(i)}) + \alpha_{GF},
$$

(14.5)

and we note that $(1 - a_{t}^{(i)})$ represents the proportion of income reported and $1 + \kappa_{GF}^{(i)}$ represents the multiplicative magnitude of the increase. We focused on the first, second and third quantiles of the multiplicative magnitudes shown in table 14.9 and together with the proportion of the reported income we took their natural logarithms. Based on equation (14.5) we then estimated the values of $\gamma_{GF}$ and $\alpha_{GF}$ and their ranges using linear regressions. Tables 14.10 shows the summary statistics
Table 14.10: Linear regression of $\log(1 + \kappa^{(i)}_{GF}) \sim \log(1 - a^{(i)}_t)$ using the results shown in Table 14.9. There was no need to run a regression for the case of the 1st quantile of the multiplicative magnitude of the increase. This is because it was equal to one irrespective of the proportion of income reported and consequently there was no Gambler’s Fallacy Effect.

of these regressions and the parameter values.

14.7 The Bomb Crater Effect

Question 10a and 10b of our survey were used to describe the Bomb Crater Effect. Respondents were asked to consider the situation where the IRS had audited them in the previous year and asked to estimate the risk of being audited again in the present year. This situation represented a hypothetical situation for most respondents. The question first asked whether they thought the risk was larger, same or smaller than their baseline audit rate. They were then asked to estimate the new audit probability. Table 14.11 summarize the results we found by providing summary statistics of the new audit probability expressed as a proportion of their baseline value. We refer to this proportion as the Bomb Crater multiplicative magnitude. Most respondents thought that the risk of being audited again in the subsequent year after an audit increased or stayed the same. This result is in direct contrast to the notion of the Bomb Crater effect where taxpayers have been found to report less income in the year that follows a tax audit. Focusing on the median values, we see that the proportion ranges from 50% for those who do believe in a decrease in the risk of being audited again, to 220% for those who believe the risk is larger. We chose to use these numbers to characterize the range of values to sample $\kappa_{BC}$. However, as shown in table 12.6, $\kappa_{BC}$ is sampled using a PERT distribution. Thus, we had to select a mode value within this range to sample. We chose to use the literature to estimate the mode value for $\kappa_{BC}$. In a laboratory experiment

<table>
<thead>
<tr>
<th>Min.</th>
<th>1st Qu.</th>
<th>Median</th>
<th>3rd Qu.</th>
<th>Max.</th>
<th>Mean</th>
<th>NA's</th>
<th>N</th>
<th>N.prop</th>
</tr>
</thead>
<tbody>
<tr>
<td>higher</td>
<td>0.00</td>
<td>1.50</td>
<td>2.20</td>
<td>4.80</td>
<td>120000</td>
<td>0</td>
<td>335</td>
<td>0.33</td>
</tr>
<tr>
<td>same</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1</td>
<td>1</td>
<td>432</td>
<td>0.43</td>
</tr>
<tr>
<td>lower</td>
<td>0.00</td>
<td>0.28</td>
<td>0.50</td>
<td>0.67</td>
<td>1</td>
<td>0.47</td>
<td>241</td>
<td>0.24</td>
</tr>
</tbody>
</table>

Table 14.11: Summary of question 10a and 10b for all records. The values in the columns ranging from the Min. to the Mean represent the new audit probability expressed as a proportion of the respondent’s baseline audit probability.

study by Kirchler, Maciejovsky, and Schwarzenberger, it was found that that compliance decreases immediately after an audit [91] to 58% when the subject taxpayers experienced a 50% effective

53Regressions were ran in R using the texreg package for display outputs [99].
tax rate, 50% penalty rate and a 30% audit risk \[91\]. Therefore, we chose to set the mode value for \(\kappa_{\text{IC}}\) at 58%. Hence, although we used the ALP to inform the range of values to sample our Bomb Crater multiplicative magnitude, we chose to go against the mean effect found by this question in the ALP survey and instead used the literature to inform the most likely value to use in our sampling scheme.

## 15 Using the ALP Survey to inform the network of social interactions

Section \[11.2\] described how the ALP survey results were used to inform the network parameters in our model. Specifically, Table \[11.7\] summarized the main inputs informing the network. In this section, we first provide a summary of the social network related questions of the ALP survey and describe the analyses that were used. We then describe in more detail the method used to assign the self-employment attribute to our taxpayers on the network using the survey results. In this and the next section, the word respondents refer to the respondents of our ALP survey and the word taxpayers refer to the agents of taxpayers considered in our ABM that interact over a social network. Thus, here taxpayers do not refer to the respondents of our survey.

### 15.1 Summary of the respondent’s social interactions

In our survey, respondents were asked to list up to ten adults that they knew, other than spouses or domestic partners, and that they interacted with on a regular basis. We call these people the *alters* of the respondent. We further refer to the set of alters that respondents talked about taxes within the past five years as the *alters of interest*. Our network related questions regarding the alters allowed respondents to select response categories labeled other, don’t know, don’t remember, preferred not to say or simply allowed the respondents to skip the question entirely. Here, we refer to these responses collectively as the other category. Most respondents listed ten alters and the average number of alters per respondent was 9.33. On average respondents had 2.63 alters of interest. However, about a third of our respondents had no alters of interest. Half of the respondents had two or less alters of interest, and 75% had four or less alters of interest. Out of the sample of respondents that had at least one alter of interest, we found their distribution of how frequently they talk about taxes. Averaging over all the respondents we found that 44% of the all tax-related interactions with the alters of interest occurred once a year, 14% occurred less frequently, 30% occurred more frequently and the remaining 12% were other responses. Due to the pronounced peak at a yearly frequency of tax-related interactions, and for the purpose of simplifying the complexity of our simulation model, we chose to consider a network of taxpayers where the edges connecting them were unweighted and described tax related interactions\[54\]. This means that each ego-alter pair in our simulation model discusses taxes once a year. Therefore, our network of taxpayers considered an unweighted non-directional network with mean degree equal to 2.63. However, the sampling range we chose to consider for the mean degree varied between 2 and 4 corresponding to the median and third quartile of the distribution of the number of alters of interest\[55\].

\[54\] The alternative would have been to consider a network of taxpayers with weighted edges and where the weights would represent the frequency of tax-related interactions.

\[55\] We chose the median instead of the first quartile for our lower bound of the sampling range for the following reason. The distribution of the number of alters of interest as reported by our respondents had a high peak at zero and the first quartile was equal to 1. Using a mean degree of zero or 1 would mean that the system includes no, or nearly no network effects. Using a mean degree equal to 1 produces a highly disconnected network containing many isolate dipoles which effectively also shuts down network effects. In this context, the critical number for a connected network
The distribution of the primary relationships of the alters with the respondents were 30% family members, 48% friends, 17% coworkers and the remaining were other responses. Instead, the distribution of the primary relationships of the alters of interest with the respondents were 48% family members, 37% friends and 12% coworkers and the remaining were other responses. Therefore, unsurprisingly alters of interest have closer relationships with the ego than alters in general. However, we did not use the primary relationship information from the survey to inform our network of taxpayers. This is because the NDSSL data for Portland, OR that we used to inform our network does not allow us to determine the type of relationship (with the exception of household members). Thus, applying these findings of the primary relationship information from the survey data to the Portland data is a difficult task and left for future work.

The distribution of the education level of the alters were 5% had no high school diploma, 29% had a high school diploma, 13% had an associate degree, 28% had a bachelor’s degree, 16% had a graduate degree and the remaining were other responses. These percentages did not change significantly when we considered just the alters of interest. As will be explained later in this section, education level of the alters was used to impute the income level of the alters, which was considered as a covariate in a regression model to help assign the self-employed attribute of the taxpayers on our representative network. The distribution of the employment type of the alters were 20% self-employed, 70% that were not self-employed. The latter included those that work for a wage and the non-working. The remaining 10% were other responses. The distribution of the employment type of the alters of interest were 29% self-employed, 69% not self-employed and 2% were other responses. These findings agree with what we intuitively expected and that the self-employed rate is higher among alters of interest. The percentage of alters that the respondent thought or knew was audited by the IRS in the past was 2.8%. This percentage increased to 4.9% for the alter of interest. For both the alters and the alters of interest who were audited, 42% were self-employed.

Out of 1030 sample of respondents, 592 were currently working and of these 179 were self-employed. The average number of alters and the average number of alters of interest of the self-employed were 9.4 and 3.2 respectively. The average number of alters and the average number of alters of interest of the non-self-employed were 9.6 and 2.5 respectively. Thus the self-employed named slightly fewer alters but had 0.7 more alters of interest, on average. The percentage of self-employed alters of the self-employed respondents was 34% and the percentage of self-employed alters of interest of the self-employed respondents was 43%. Likewise, the percentage of non-self-employed alters of the non-self-employed respondents was 76% and the percentage of non-self-employed alters of interest of the non-self-employed respondents was 74%. Thus, using these numbers we found that the average number of self-employed alters of the self-employed respondents is 3.2 (i.e., 9.4 · 0.34) and the average number of self-employed alters of interest of the self-employed respondents is 1.4 (i.e., 3.2 · 0.43). Thus, 44% (i.e., 1.4/3.2) of all relationships involving the self-employed included talking about taxes. Likewise, we used a similar approach to calculate the remaining three other proportions between the self-employed and non-self-employed relationships shown in Table 11.7 in Section 11.2.

Out of the 179 self-employed respondents, 130 were never audited and 49 were audited. The mean number of alters of interest of these groups were respectively 3.1 and 4. Thus, from these
numbers for the self-employed, the multiplicative effect representing how many more relationships those who were audited have compared to those who were not audited is 1.29 (i.e., 4/3.1). Out of the 413 (i.e., 592-179) non-self-employed respondents, 348 were never audited and 65 were audited. The mean number of alters of interest of these groups were respectively 2.4 and 3.0. Thus, from these numbers for the non-self-employed, the multiplicative effect representing how many more relationships those who were audited have compared to those who were not is 1.25 (i.e., 3/2.4). A similar number is shown in Table 11.7 in Section 11.2.

We found that the mean and median of the perceived audit rate distribution of our full set of respondents did not vary based on whether respondents reported to have ever been audited or not or to know of anyone in their social network to have ever been audited. Statistical analysis showed that there was no significant difference between the mean perceived audit rate for these different groups.

15.2 Predicting and assigning the self-employment attribute

We used our ALP survey data to build a predictive model of self-employment, which we applied to the taxpayers in our ABM. This is because self-employment is not an available attribute in the NDSSL data for Portland OR, that we used to inform the taxpayer’s social network. The aim of our predictive model is to assign the self-employment attribute to our taxpayers on the network and reclaim the assortative mixing patterns observed in the survey data. Here we described our approach to this task. We note that the purpose of the predictive model is not to present new empirical results. Therefore, we do not describe and interpret the results of the regressions describing our predictive model.

We used the social network questions of our ALP survey described in Section 13 to run separate logistic regression models to predict whether a respondent is self-employed based on the covariates for (i) age, (ii) gender, (iii) income (in units of $100,000), (iv) average income of the alters (in units of $100,000) and (v) the proportion of alters that are self-employed. As mentioned, the purpose of these regressions was to use the ALP data to inform who in our network of taxpayers is self-employed and who is not. Since our survey did not provide the actual income of the respondents nor of their alters, we have had to impute them. To obtain the income of the respondent we combined the categorical responses to the familyincome and familyincome_part2 questions into one variable. Then we used the CPS 2015 data mentioned in Section 11.2 and randomly assigned an income based on both the combined income categorical variable and the CPS income distribution. To impute the income level of the alters of the respondents we used the Bayesian regression model by Liu et al. [107]. We used this regression to impute the income level of the alters of each respondent in our survey based on the alter’s level of education and gender.

To predict the employment status, our first logistic regression considered just the first three covariates, namely age, gender, and income. The second regression considered the first four covariates (i.e., including the average income of the alters) and the final regression considered all five covariates. Table 15.1 provides the summaries of these regression models. From our logis-

56These numbers imply cumulative audit rates ranging from 65/413 to 49/179, or equivalently ranging between 16% and 27%. This may appear to be very large. However, assuming 30 years of work and a 0.8% annual audit rate (for all workers) gives a cumulative probability of 21%, which is bracketed by the two numbers above. Note that 30 years of work is justified based on the average age of our ALP respondents.

57We also examined several other model specifications (not shown here). Some of these specifications had a quadratic term of the income variable. We considered the quadratic term in the model because we hypothesized a quadratic relationship between the probability of being self-employed and income. Specifically, we supposed that people who are in lower and higher income brackets might be more likely to be self-employed than those in the middle-income brackets. However, we were not able to observe such relationship in our data. In some other specifications, we considered in-
Table 15.1: Regression Models predicting self-employment status.

<table>
<thead>
<tr>
<th></th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>-3.88***</td>
<td>-3.63***</td>
<td>-4.02***</td>
</tr>
<tr>
<td></td>
<td>(0.49)</td>
<td>(0.54)</td>
<td>(0.58)</td>
</tr>
<tr>
<td>genderMales</td>
<td>0.40*</td>
<td>0.35</td>
<td>0.28</td>
</tr>
<tr>
<td></td>
<td>(0.19)</td>
<td>(0.19)</td>
<td>(0.20)</td>
</tr>
<tr>
<td>age</td>
<td>0.05***</td>
<td>0.05***</td>
<td>0.05***</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
</tr>
<tr>
<td>income.val.5</td>
<td>0.09</td>
<td>0.14</td>
<td>0.10</td>
</tr>
<tr>
<td></td>
<td>(0.12)</td>
<td>(0.13)</td>
<td>(0.14)</td>
</tr>
<tr>
<td>alter_income.val.5</td>
<td>-0.77</td>
<td>-1.43</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.85)</td>
<td>(0.92)</td>
<td></td>
</tr>
<tr>
<td>alter_se_proportion</td>
<td>3.38***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.44)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AIC</td>
<td>675.78</td>
<td>667.73</td>
<td>604.22</td>
</tr>
<tr>
<td>BIC</td>
<td>693.28</td>
<td>689.53</td>
<td>630.38</td>
</tr>
<tr>
<td>Log Likelihood</td>
<td>-333.89</td>
<td>-328.86</td>
<td>-296.11</td>
</tr>
<tr>
<td>Deviance</td>
<td>667.78</td>
<td>657.73</td>
<td>592.22</td>
</tr>
<tr>
<td>Num. obs.</td>
<td>587</td>
<td>579</td>
<td>579</td>
</tr>
</tbody>
</table>

**p < 0.001, *p < 0.01, *p < 0.05**

Table 15.2: Consolidated Regression Models predicting self-employment status.

<table>
<thead>
<tr>
<th></th>
<th>Model 1</th>
<th>Model 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>-3.85***</td>
<td>-4.34***</td>
</tr>
<tr>
<td></td>
<td>(0.49)</td>
<td>(0.53)</td>
</tr>
<tr>
<td>genderMales</td>
<td>0.41*</td>
<td>0.32</td>
</tr>
<tr>
<td></td>
<td>(0.19)</td>
<td>(0.20)</td>
</tr>
<tr>
<td>age</td>
<td>0.05***</td>
<td>0.05***</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.01)</td>
</tr>
<tr>
<td>alter_se_proportion</td>
<td>3.16***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.42)</td>
<td></td>
</tr>
<tr>
<td>AIC</td>
<td>680.20</td>
<td>620.21</td>
</tr>
<tr>
<td>BIC</td>
<td>693.34</td>
<td>637.73</td>
</tr>
<tr>
<td>Log Likelihood</td>
<td>-337.10</td>
<td>-306.11</td>
</tr>
<tr>
<td>Deviance</td>
<td>667.78</td>
<td>657.73</td>
</tr>
<tr>
<td>Num. obs.</td>
<td>587</td>
<td>579</td>
</tr>
</tbody>
</table>

**p < 0.001, *p < 0.01, *p < 0.05**

tic regressions shown in table [15.1] only age, gender and the proportion of self-employed alters are significant in predicting self-employed status of the ego. Therefore, we considered two reduced regression models. The first includes age and gender and the second also includes the proportion of self employed alters. Table [15.2] shows the summary statistics of these two regression models. None of these models provided a significant evidence for a linear or curve-linear relationship between the probability of self-employment and income.
We used these regression models to assign the self-employment attribute on our network using the following iterative method. For each taxpayer on our network we used the first regression to predict an initial estimate of the probability of being a self-employed worker. Here, we denote this probability by $P_{1}^{(i)}$ for taxpayer $i$ and it is given by

$$\text{logit}(P_{1}^{(i)}) = \log \left( \frac{P_{1}^{(i)}}{1 - P_{1}^{(i)}} \right) = \kappa^{(1)} + \beta_{\text{gender}}^{(1)} \cdot \text{gender} + \beta_{\text{age}}^{(1)} \cdot \text{age}, \quad (15.1)$$

where $\kappa^{(1)}$, $\beta_{\text{gender}}^{(1)}$, $\beta_{\text{age}}^{(1)}$ are respectively the intercept and the regression coefficients of the first regression model. The subscript in $P_{1}^{(i)}$ denotes the first iteration of the process. These probabilities were used to randomly assign the self-employment attribute to each taxpayer on the network for our first iteration. It further allowed us to calculate the initial expected proportion of alters that are self-employed for each taxpayer. We denote this attribute as $\langle P_{1}^{(i)} \rangle_{nn}$, where the first subscript denotes the first iteration and the second subscript $nn$ denotes that the average is over the nearest neighbors (i.e., the alters) of $i$. In our second iteration, we used the second regression model that includes the proportion of self-employed alters as a covariate in order to refine our estimate of the probability of being a self-employed worker. We denote this probability by $P_{2}^{(i)}$ for taxpayer $i$ and it is given by

$$\text{logit}(P_{2}^{(i)}) = \kappa^{(2)} + \beta_{\text{gender}}^{(2)} \cdot \text{gender} + \beta_{\text{age}}^{(2)} \cdot \text{age} + \beta_{\text{se}}^{(2)} \langle P_{1}^{(i)} \rangle_{nn}, \quad (15.2)$$

where $\kappa^{(2)}$, $\beta_{\text{gender}}^{(2)}$, $\beta_{\text{age}}^{(2)}$ and $\beta_{\text{se}}^{(2)}$ are the coefficients of the second regression model. The subscript in $P_{2}^{(i)}$ denotes the second iteration of the process. We used this second regression which factors in the effect of the proportion of alters that are self-employed to randomly reassign the self-employment attribute. For each taxpayer on our network, we then recalculated the proportion of alters that are self-employed. The second regression model was used to perform a few more iterations of the process by

$$\text{logit}(P_{k+1}^{(i)}) = \kappa^{(2)} + \beta_{\text{gender}}^{(2)} \cdot \text{gender} + \beta_{\text{age}}^{(2)} \cdot \text{age} + \beta_{\text{se}}^{(2)} \langle P_{k}^{(i)} \rangle_{nn}, \quad (15.3)$$

where the subscript $k$ denotes the iteration. In each iteration we reassigned the self-employment attribute for each taxpayer and found the new proportion of alters of each taxpayer that are self-employed. This procedure converged rapidly and only a few iterations are needed to reach a point where the self-employment attribute of the taxpayers no longer changes. Due to this rapid convergence, we chose to carry out just four iterations. Specifically, $P_{4}^{(i)}$ was used to inform the self.employment.propensity attribute and was used by our model to sample the self-employed taxpayers on our network and thus informing the self.employment attribute. Our target proportion of self-employed was $16.8\%$ \footnote{This target was informed by IRS 2016 statistics where there were 24 million self employed filers and 119 million non self-employed filers. However, this accounts for tax filers. According to the 2016 U.S. Bureau Of Labor Statistics, the true proportion of self-employed should be around $10\%$ \footnote{We used the numbers from the IRS tables to get the proportion of self-employed.}. This procedure allowed us to match this target: the proportion of self-employed taxpayers in our population is $16.82\%$.}

This target was informed by IRS 2016 statistics where there were 24 million self employed filers and 119 million non self-employed filers. However, this accounts for tax filers. According to the 2016 U.S. Bureau Of Labor Statistics, the true proportion of self-employed should be around $10\%$ \footnote{We used the numbers from the IRS tables to get the proportion of self-employed.}.
16 Using the ALP Survey to inform the taxpayer’s behavioral attributes

Previously, in Section 11.1 we described the taxpayer’s static attributes. These are listed in Table 11.2. Some of the attributes were informed by the ALP survey. These include the initial/baseline perceived audit rate ($\tilde{q}^{(i)}_0$) and penalty rate ($\tilde{P}^{(i)}_0$), the value for $c^{(i)}_1$ and the actor.logical. In this section, we describe how the values for these static attributes were assigned to our taxpayers on the network in our ABM using the results from our ALP survey. Using the taxpayer’s perceived audit rate attribute as our example, we first describe a general method that we used repeatedly to assign values of the other static attributes of our taxpayers. This method relies on using regression models. As mentioned in the previous section, the purpose of these regressions is not to present new empirical results. Therefore, we do not describe and interpret the results of these regressions. A more detailed empirical regression-based analyses of our data will be described in a different report.

16.1 Assigning the taxpayer’s perceived audit and penalty rates

The initial/baseline perceived audit ($\tilde{q}^{(i)}_0$) of the taxpayer population in our model was informed using a general method that relies on two statistics. The first used a linear regression model of the survey data that predicts the perceived audit rate from the relevant covariates using the responses of the entire sample. The second was the distribution of the perceived audit rate of a selected sample of ALP survey respondents. This sample included respondents that had never experienced an IRS audit and that did not know of any of their close contacts having ever being audited, including their spouse at the time of the survey. Our linear regression analysis considered the dependence of the perceived audit rate on a large number of covariates. The main set of covariates included age, gender, income (measured in units of $100,000), perceived effective tax rate, whether the respondent was ever audited, whether the respondent’s spouse was ever audited, whether the respondent is self-employed, the proportion self-employed alters, the perceived evasion rate among one hundred people like themselves, the perceived evasion rate in the population, the level of satisfaction that the services they receive are worth the taxes they pay and the famous actor effect.\footnote{The response for the famous actor effect were recoded so that a value of 5 indicated that the respondent would be more likely to fully report their income and a value of 1 indicated much less likely to fully report their income.} However, for some respondents in the dataset, a few of these covariates are missing. This is because either the respondent did not choose to answer the question or the question was not applicable. For example, the question on the self-employment status was not applicable to respondents that self-identified as not employed. Likewise, not all respondents had a spouse or were in a serious relationship. In order to use all the available data in our regression analysis, covariates with missing values were overwritten with means/neutral values found from the corresponding distribution across the respondents. For example, for a respondent with no spouse or not in a serious relationship, the survey question asking whether their spouse was ever audited was skipped and resulted in a missing value. This missing value was then replaced by a number between zero and one given by the mean response to this question. The mean was calculated across the sample of respondents that received and answered this question. Then, for each covariate, we constructed a vector of indicator variables where the vector elements recorded whether or not a respondent had a missing value that was overwritten by this process. Thus, in our example, we used an indicator variable to record all those respondents without a spouse that had their entry replaced by the mean value. This vector of indicator variables was used as an additional covariate in the regression model to help identify whether the problem of missing entries...
Refined Model

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>14.25</td>
<td>(2.47)</td>
<td>***</td>
</tr>
<tr>
<td>calcage</td>
<td>-0.08</td>
<td>(0.03)</td>
<td>*</td>
</tr>
<tr>
<td>genderFemale</td>
<td>7.83</td>
<td>(0.92)</td>
<td>***</td>
</tr>
<tr>
<td>income.val.5</td>
<td>-4.84</td>
<td>(0.67)</td>
<td>***</td>
</tr>
<tr>
<td>perceivedtaxrate</td>
<td>0.19</td>
<td>(0.03)</td>
<td>***</td>
</tr>
<tr>
<td>prop.alters.Taxes</td>
<td>0.47</td>
<td>(0.19)</td>
<td>*</td>
</tr>
<tr>
<td>altertaxaudit.Taxes</td>
<td>10.09</td>
<td>(3.70)</td>
<td>**</td>
</tr>
<tr>
<td>selfemployedTRUE</td>
<td>-5.19</td>
<td>(1.21)</td>
<td>***</td>
</tr>
<tr>
<td>actor</td>
<td>3.38</td>
<td>(0.61)</td>
<td>***</td>
</tr>
<tr>
<td>perceivedevasionrate</td>
<td>0.12</td>
<td>(0.02)</td>
<td>***</td>
</tr>
<tr>
<td>actor__NA.IndicatorTRUE</td>
<td>-1.17</td>
<td>(2.33)</td>
<td></td>
</tr>
</tbody>
</table>

**AIC**: 8424.02  
**BIC**: 8483.27  
**Log Likelihood**: -4200.01  
**Deviance**: 209986.19  
**Num. obs.**: 1030

* *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Table 16.1: Regression Models predicting perceived audit rate. This refined linear regression model considers only the covariates which were found to be significant with a p-value below 5% from a previous linear regression model that considered a fuller set of covariates. We see that although the response to the actor variable is significant, we need to take into account that missing responses to this question also contributed to the prediction model.

for a given covariate had a significant impact on the prediction. Table 16.1 shows the results of the linear regression model for the perceived audit rate. We used the regression model shown in Table 16.1 to predict the first set of perceived audit rates for our taxpayers. The prediction relied on the set of available attributes of our taxpayer population. Thus, not all covariates could be used in the prediction. For example, our taxpayer’s population did not have the perceived evasion rate values and thus the covariate value that entered the prediction was taken from the mean value across the respondents of the survey. We used this first set of predicted perceived audit rates to find the ranking and identify which taxpayer should be assigned the highest perceived audit rate, which the second highest and so forth.

The actual values for the perceived audit rate were instead assigned to our taxpayers by sampling from a fitted smooth distribution of the perceived audit rates taken from our ALP survey. Figure 16.1 shows the histogram of the perceived audit rate in our selected sample of respondents and a smoothed curve that describes the distribution. The figure shows the distribution and the smoothed curve fit on both a log-linear and a log-log scale. The smooth curve showed in Figure 16.1 was used to sample our set of perceived audit rates. The median and mean perceived audit rate of our sample were 15% and 18% respectively. Some respondents in this selected sample gave very large perceived audit rates. For example, 15% of our sample gave a perceived audit rate above 34%, 10% gave a perceived audit rate above 40% and 0.5% gave a perceived audit rate above 80%. We chose to bound our sampling of the perceived audit rate to below 80%, and hence this is a bias. This only affected about 0.5% of the sample. The reason we chose to bound our perceived audit rate to 80% for these taxpayers was to allow them to have some margin to increase their per-

60In this report we do not describe and interpret the results of the regression models. This will be done in another publication. Instead, we use these regression models as a statistical guidance in order to assign baseline attribute values to our taxpayer population.
Figure 16.1: Perceived audit rate distribution amongst the respondent that never experienced an IRS audit and that did not know of any of their close contacts having ever being audited. Both plots show the perceived audit rate expressed as a percentage and in a log 10 scale. Plot (b) shows the counts in a log 10 scale. The real audit rate is roughly 0.8% and although some respondents did provide perceived audit rates that were lower than 1%, they were surprisingly very few. Using the log 10 scale for counts, plot (b) allows us to get a better sense of how many respondents thought that the audit rate was lower than 1%. The blue curve shows the smoothed fit to the distribution. We used the smoothed fit to the distribution shown in plot (a) to sample our perceived audit rates for the model.

This same method was used to assign the taxpayer’s perceived penalty rates. Table 16.2 shows the results of the regression model. Figure 16.2 shows the smoothed curve used for the sampling the distribution of the perceived penalty rates.

The goal of this method was to provide a reasonable set of values to inform our taxpayer’s perceived audit and penalty rates used by the ABM. Admittedly, the estimates produced by this approach are approximate. The value of added precision in these estimates is limited by recognizing that the many sources of uncertainty in the other parameter values that inform the ABM, that have been described in Section 12, have larger effects due to the high sensitivity of model out-

<table>
<thead>
<tr>
<th>Refined Model</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>76.01 (12.54)***</td>
</tr>
<tr>
<td>calcAge</td>
<td>-0.81 (0.20)***</td>
</tr>
<tr>
<td>prop. alters. tTaxes</td>
<td>2.25 (1.18)</td>
</tr>
<tr>
<td>AIC</td>
<td>12168.53</td>
</tr>
<tr>
<td>BIC</td>
<td>12188.28</td>
</tr>
<tr>
<td>Log Likelihood</td>
<td>-6080.27</td>
</tr>
<tr>
<td>Deviance</td>
<td>8087088.10</td>
</tr>
<tr>
<td>Num. obs.</td>
<td>1030</td>
</tr>
</tbody>
</table>

Table 16.2: Regression Models predicting perceived penalty rate
puts to the choices of these values. Therefore, for the purpose of the ABM, this method provides a good enough approximation which reproduces the distribution of perceived audit and penalty rates found in the ALP while assigning also these values according to general dependencies found by these regression models. An alternative method could have used a generalized linear regression model to directly predict and assign the taxpayer’s perceived audit rate. A generalized linear regression model requires that we specify a link function that is used to describe the relationship between the linear predictor and the mean of the distribution function. The choice of the link function affects the prediction model, and although there are many commonly used link functions it is unclear which one is best to use in our case. In our approach, instead of specifying one of the commonly used link functions we essentially use the empirical distribution to act as our link function.

### 16.2 Assigning the taxpayer’s $c_1$ values

In Section 14.2.4 we described how we generated the values of $c_1$ for the survey respondents. Figure 14.6 showed two distributions that we separately used to sample the $c_1$ values for the taxpayers for our simulation model. The assignment of the sampled $c_1$ values using each of the two distributions to the taxpayers in our ABM followed the same method as the one described for the perceived audit and penalty rates. We first ran a linear regression model of our $c_1$ values on the same set of covariates described previously. We then used the ranking of the predicted $c_1$ values from our regression model together with the two distributions for $c_1$ shown in Figure 14.6 to assign our lower and upper bound estimates of $c_1$ for each of our taxpayers. In other words, the $c_1^{(i)}$ values predicted from the regression were used for ranking purposes only and thus to identify which taxpayer should be assigned the highest $c_1^{(i)}$ value, the next to highest value and so forth from our sampled set. Therefore, each taxpayer was assigned two different $c_1$ values taken from the distributions shown in Figure 14.6. However, the order of the values of the assigned $c_1$ across the taxpayers in our population was the same as that informed by the regression model.

Table 16.3 shows the results of the regression. We note that both the perceived tax rate and perceived evasion rate are significant predictors in the regression model. This is not surprising since the value of $c_1$ was derived using the gradient and parallel transportation method on a plot of per-
Table 16.3: Regression model predicting the values of $c_1$.

<table>
<thead>
<tr>
<th></th>
<th>Refined Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>35.66 (2.30)***</td>
</tr>
<tr>
<td>perceivedtaxrate</td>
<td>0.29 (0.07)***</td>
</tr>
<tr>
<td>servicestaxes</td>
<td>2.32 (0.92)∗</td>
</tr>
<tr>
<td>perceivedevasionrate</td>
<td>−0.39 (0.05)***</td>
</tr>
<tr>
<td>perceivedtaxrate__NA.IndicatorTRUE</td>
<td>4.68 (6.23)</td>
</tr>
<tr>
<td>AIC</td>
<td>10135.34</td>
</tr>
<tr>
<td>BIC</td>
<td>10164.96</td>
</tr>
<tr>
<td>Log Likelihood</td>
<td>-5061.67</td>
</tr>
<tr>
<td>Deviance</td>
<td>1118969.77</td>
</tr>
<tr>
<td>Num. obs.</td>
<td>1030</td>
</tr>
</tbody>
</table>

∗∗∗$p < 0.001$, ∗∗$p < 0.01$, ∗$p < 0.05$

received tax rate versus perceived evasion rate as described in Section 14.1.3. Thus, this dependence is inbuilt in our approach and expected. Thus, these regression results act as a verification check. Interestingly, the regressions reveal that $c_1$ also depends on the covariate servicestaxes informed by question 32 in our ALP survey. This describes the perceived extent that public goods and services received by the respondents are worth the federal income taxes they pay. Thus, a respondent’s estimated $c_1$ is greater, and hence s/he is more compliant if s/he perceives that the services received in return for his/her paid taxes are worth it.

Although our regression analysis of the ALP survey data found that there are three covariates that are significant in predicting the $c_1$, the only covariate used to make our prediction was the perceivedtaxrate. Moreover, we used the effective tax rate applied to the taxpayer, and calculated based on his/her income and filing status, as our perceivedtaxrate covariate input to the regression. The other two significant covariates that appear in our regression analysis include servicestaxes and perceivedevasionrate. These could not be used in the prediction since our taxpayers do not have these attributes. Consequently, these covariates have a neutral effect in predicting the ranking order of the dependence of the $c_1$ values on the effective tax rate. The reason we report $c_1$’s dependence on perceivedevasionrate in the regression is to show that the way respondents think of concepts relating to their own compliance behavior, such as $c_1$, depends on their perception of compliance levels of others. Therefore, we show this for the purpose of further justifying assumptions made in the conceptualization of our model. The reason we report $c_1$’s dependence on servicestaxes, is to suggest that future extensions of our ABM should include a dynamic description of how the perception of the cost to quality of public services changes over time and affects the taxpayers’ compliance behavior.

16.3 Actor effect

We have previously described the actor/famous person effect. The taxpayer’s actor.logical attribute is an indicator variable that determines whether a taxpayer is influenced by the actor/famous person effect. The assignment of this attribute to the taxpayers was based on a logistic regression model of our survey respondents. The dependent variable in our regression was constructed based on the responses to our survey question 43. This produced an indicator variable for each

---

61In our model, taxpayers are affected by changes in tax morale of their alters and by what is reported by the media in terms of compliance levels. These are strongly correlated to social network and population-level compliance behaviors. Hence, although we do not explicitly track a dynamic attribute for perceivedevasionrate for our taxpayers, this is implicitly captured the way we model social and media influences.
respondent which was equal to one if the respondent was more or much more likely to fully report their income if s/he heard that a famous actor was caught and prosecuted for tax evasion. Again,

<table>
<thead>
<tr>
<th></th>
<th>Refined Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>-0.63 (0.39)</td>
</tr>
<tr>
<td>calage</td>
<td>-0.02 (0.01)**</td>
</tr>
<tr>
<td>income.val.5</td>
<td>-0.76 (0.18)**</td>
</tr>
<tr>
<td>perceivedtaxrate</td>
<td>0.02 (0.01)**</td>
</tr>
<tr>
<td>perceivedauditrate</td>
<td>0.03 (0.00)**</td>
</tr>
<tr>
<td>AIC</td>
<td>938.09</td>
</tr>
<tr>
<td>BIC</td>
<td>962.78</td>
</tr>
<tr>
<td>Log Likelihood</td>
<td>-464.05</td>
</tr>
<tr>
<td>Deviance</td>
<td>928.09</td>
</tr>
<tr>
<td>Num. obs.</td>
<td>1030</td>
</tr>
</tbody>
</table>

Table 16.4: Logistic regression model used to predicting if a taxpayer is affected by the actor/-famous person effect.

we used the same set of covariates described previously for our regression model. Table 16.4 shows the results of our regression model using only the set of covariates which were found to be significant. Thus, in addition to age and income, also perceivedtaxrate and perceivedauditrate were used for this prediction. However, as before we used the effective tax rate applied to the taxpayer to replace perceivedtaxrate in our prediction model. Moreover, we used the predicted perceived audit rate of the our taxpayers described in Section 16.1 to make the prediction of the actor.logical attribute. Thus, in our approach the attribute actor.logical is assigned to the taxpayers after we have assigned the perceivedauditrate attribute.
Model Verification, Validation and Calibration

17 Model Verification and Validation

17.1 Differences between Model Verification and Model Validation

The first step in model testing requires that we check whether the model implements the assumptions correctly. This type of testing is known as model verification. It is limited in scope as its sole aim is to make sure that the implementation of the model code accurately reflects the model design and formulation. Model verification also checks that the outputs of single model components generally produce the intended behavior when run in isolation. Once the model is thoroughly verified we can proceed to model validation. This type of model testing checks that the model assumptions are reasonable with respect to the real system. Thus, we validate the model by comparing the model output to what is generally known about the real system and check whether the model can reproduce known "stylized facts". Model verification does not imply validation, nor validation implies verification. However, when measurement data is available for the system being modeled there is often in practice an overlap between verification and validation tests. In our case, our description of model testing provides an overlap between these two types of model testing. In what follows, we will distinguish what test are verifications and what tests are both verifications and validations. Additional model validations are described in the section on our uncertainty and sensitivity analysis.

17.2 Model component verifications

17.2.1 Summary of previously described model component verifications

Model verification was done throughout the model development. Each model component, or module, was first tested in isolation from the feedback effects of other model components. Using artificially set up environments and inputs we tested how the module responded to these inputs and whether it produced expected results, and if not what were the reasons for it. Below we provide a few examples of model verifications we have already described in previous sections and provide a brief summary.

- In section 7.1 we described how our tax morale module of our taxpayers works. Figure 7.1 shows one example of the many model verifications we carried out to test this module.

- In section 7.5.2 we described how perceived audit rates of our taxpayers change dynamically in the model. We carried out various model verifications here. For example, figure 7.3 shows how perceived audit rates of the taxpayers change with an increasing audit rate and compared this dependence to previously assumed relationships.

- In section 8 we described the IRS module. In particular, we described IRS data and our assumptions informing how our model for the IRS selects taxpayers to be audited. As given in table 8.3 we also estimated the average cost per audit broken by income and by type of audit. By running this module together with the full ABM, we verified and validated that our assumptions reproduce a very similar value for the overall average cost per IRS audit.

---

62Quoting Wikipedia, "A stylized fact is often a broad generalization that summarizes data, which although essentially true may have inaccuracies in the detail."
• In section 11.2 we described how we constructed our synthetic population and the network that connects the taxpayers. There we provided a description of our verifications that tested the network statistics and reported the level of agreement we get when compared to the network data collected by our survey. For example, Figure 11.1 provides a verification of some network statistics and how we filtered our full representative data set for Portland to get smaller representations of the same overall network.

• Appendix A provides a lengthy description and analysis, including verification analyses of our implementation of the experience discounting and the EWMA process we assume in our taxpayer’s behavioral model.

Figure 17.1: Network plots used to verify the structure of the network of taxpayers and its effect on the taxpayer compliance described by the model dynamics. Figure (a) illustrates the whole PN1 network using a similar display to the network plot shown in Figure 11.2 but with a different orientation and transformation of the position of the nodes using an analytical method for squaring the disc [57]. As, before, nodes represent taxpayers and edges represent relationships over which taxpayers can influence each others tax morale and risk perceptions. Colors represent income categories from dark red to navy blue in group breaks at $25,000, $50,000, $100,000, $250,000 and $500,000. The nodes with the diamond symbol are the self-employed taxpayers. Figure (b) illustrates the same network during the model dynamics. Here, the visibility or opacity of the nodes gives an indication of the proportion of hideable income that is underreported. The opacity of the edges represents whether taxpayers that are underreporting by similar proportional amounts are connected.

17.2.2 Additional network related model verifications

Additional network verification plots are shown in Figure 17.1. These plots illustrate the whole PN1 network and how compliance changes over the entire network during the model dynamics. Specifically, Figure 17.1(b) was produced for each year of the model dynamics and was used for
model verification. However, these network plots show the entire population network and although we produced interactive network plots\textsuperscript{63}, these were still hard to use and interpret for the purpose of verifying the effects of the network dynamics. Therefore, during the model verification, we manually looked at how each taxpayer (i.e., the ego) and his/her alters interacted, focusing the dynamic attributes describing compliance behavior of the ego and how they were correlated to those of the alters. To assist us we produced some egocentric network visualizations of the dynamics. An example is shown in Figure 17.2

\begin{figure}[h!]
\centering
\includegraphics[width=\textwidth]{figure17.2.png}
\caption{Egocentric network representation of compliance behavior produced by the model dynamics and used for model verification. This illustration focuses on a self-employed taxpayer representing our egocentric node and shown by the large red central diamond-shaped node in the middle. For the purpose of showing an interesting case, we chose a self-employed taxpayer with as many as five alters with which s/he talks about taxes. This is more than the average number of alters each taxpayer has in the network. Alters, shown in the plot can be self-employed (diamond shaped nodes) or non-self-employed workers (circle shaped nodes). Nodes shown in red indicated a higher level of tax evasion of their hideable income. The level of tax evasion decreases as we go from red-colored nodes to bright yellow colored nodes. Edges are also colored depending on the compliance level of the nodes they connect. In an interactive version of these plots, we show dynamic tax compliance behavioral attributes besides each node. This interactive visualization assisted us with our model verification.}
\end{figure}

17.3 Model verifications using first versions of our model implementation

The conceptualization of our taxpayers behavioral model strongly suggested that the dynamics of our model would lead to a segregation whereby taxpayers get polarized into either full tax compliance behavior or full tax evasion behavior depending on his/her tax morale and risk perceptions. Hence, producing a U-shaped distribution of the proportion of hideable income reported. This dynamic behavior was confirmed early on by model verifications of the very first versions of our model implementation. These simplified versions of the model were not informed by data nor by expert opinion and made many simplifying assumptions. For example, here we assumed a homogenous population whereby the income for all taxpayer was fully hideable and all taxpayers had 100 units of income per year and were taxed at an effective tax rate of 30%. The only source

\textsuperscript{63}We used the R package called visNetwork to create interactive versions of these plots.
of heterogeneity was the values of $c_1^{(i)}$ for each taxpayer. Figure 17.3 shows plots from these first versions of our model. Nevertheless, even our simplest model implementation produced a clear segregation in tax compliance behavior and a U-shaped distribution.

Figure 17.3: First model verification using the conceptualized model implementation and test inputs that were not informed by data and for a homogenous population where all taxpayers has an income of 100 units that was considered completely hideable. Plot (A) shows the sample model dynamics of aggregated variables. In particular the tax gap reaches very large values of over 60%. Plot (B) shows how the population self-segregated and polarized into two different tax compliance behaviors.

Figure 17.4: Empirical distribution of reporting compliance rates for Taxpayer Sample using NRP data for TY 2006 to TY 2009 and the sampling weights as reported in Alm et al. [6]. The orange section of the histogram shows provides a sense of the proportion of the "pathologically honest" taxpayers who report over 100% of their taxable income.

In a 2015 paper by Alm, Bloomquist and McKee [6], the authors conducted laboratory experiments on tax compliance, and compared tax compliance behaviors in the laboratory to those found from a large sample of actual taxpayers. They showed that both the taxpayer sample and
the experimental sample have a bimodal highly polarized U-shaped distribution. The U-shaped
distribution of the taxpayer sample compliance behaviors using the sample weights is shown in
Figure 17.4. Thus, the U-shaped distribution produced by our model reveals a validation of our
model as it is able to reproduce this stylized fact of tax behavior compliance. The actual shape and
proportion of taxpayers that remain fully compliant in the laboratory experiments was shown to
depend on the audit rate used. Thus, the higher the audit rate the higher the compliance. How-
ever, the audit rates in the laboratory experiments ranged between 0% and 30% which represents
a very large range. Our U-shaped distribution also showed a strong dependence on the audit rate
within such a range.

17.4 Model verifications and validations using our calibrated model run

In contrast to the first versions of our model implementation, our full model was informed by all
data sets described in this report including IRS data, survey data, and network data. Therefore,
our full model considers a network and synthetic population where taxpayers attributes are het-
erogeneous. Throughout the development of the full model, we carried out multiple verifications.
Here, we illustrate some results produced by these verifications. The model starts off with an
un-natural arbitrary initial condition whereby all taxpayers are initialized as being compliant and
have no tax compliance and IRS enforcement experience. The model is then run until the system
reaches a state of dynamic equilibrium whereby the average tax gap over a time window of 40
years, and its deviations about the average is stationary. The particular model settings used
for this illustration is based on one of our model cases runs taken from a set of 2,500 case runs.
The input parameter values for each of these case runs was generated using the Latin Hypercube
Sampling approach described in section 12.1. More precisely, the case run we illustrate belongs to
a calibrated subset of the initial 2,500 that can reproduce our output model targets within a given
tolerance.

17.4.1 Verification of the taxpayer’s compliance behavior trajectories

Figures 17.5 and 17.6 shows compliance trajectories for both self-employed and non-self-employed
working taxpayers over the course of 400 years using for our best case calibrated model. In
these plots, the dotted pink vertical lines indicate an IRS audit and vertical gray solid lines indicate
an IRS audit, detection, and penalty for tax evasion. Notice how taxpayers that are non-compliant
can become compliant when audited even when the tax-evasion is not detected. The plots also
reveal how more frequently, the self-employed are selected for an audit compared to the non-self-
employed.

---

64 Appendix I describes how we tested the dynamics to determine whether the system has reached equilibrium.
65 Model calibration is described in detail in Section 20.
66 As noted in our model description and in particular in section 5.1, our model does not account for vital dynamics.
However, the taxpayers’ generation half-life is the longest time scale in our model and is several decades long. Over
this timescale, all tax-related experiences acquired by taxpayers are lost. As such the trajectories can be interpreted as
the compliance behavior of multiple taxpayers over the years, one following after the other who are assured to have
similar static attributes but go through different tax-related experiences. Thus, a sort of family-level compliance over
the decades and centuries.
Figure 17.5: Example compliance trajectories for four self-employed taxpayers using our best case calibrated model run.

Figure 17.6: Example compliance trajectories for four non-self-employed taxpayers using our best case calibrated model run.
17.4.2 Verification of the dynamics of aggregate-level compliance behavior

Figure 17.7A shows the dynamics of the tax gap, the mean perceived audit rate and the mean perceived penalty rate. Figures 17.7B also shows the dynamics for the tax gap by providing an iterative-map representation of the dynamics and shows how the tax gap reaches a dynamic equilibrium where the tax gap fluctuates between 15.5% and 16%. Notice how the mean perceived audit rate starts from its initial value and builds up over time, increasing by as much as 6% over the course of the dynamics. Recall that the starting initial perceived audit rate values in our population are informed by the ALP survey using respondents that have never been audited and that do not know of any alters who has ever been audited.

17.4.3 Verification of the distribution of the thresholds describing tax evasion

Figure 17.8: Histograms showing the distribution of the $c_1^{(i)}$ threshold value (red), and the distribution of the $\tilde{c}_1^{(i)}$ threshold value at equilibrium of our taxpayers in the model.
Figure 17.8 shows a histogram (red) of the $c_1^{(i)}$ threshold value which is a static attribute of our taxpayers. It also shows a histogram (blue) of the equilibrium distribution of the $\tilde{c}_1^{(i)}$ threshold value which is instead a dynamic attribute of our taxpayers. Despite taxpayers have a relatively low value for $c_1^{(i)}$, when they factor in their perceived levels of deterrence, their $\tilde{c}_1^{(i)}$ becomes much larger and bell-shaped. Hence, with perceived deterrence, our compliant taxpayers have a much stronger propensity to not-initiate in tax evasion. As explained in our model description, once taxpayers do initiate in tax evasion behavior their compliance behavior in future years is largely determined by how successful they are in not reporting parts of their hideable income and not getting caught by the IRS.

### 17.4.4 Verification and validation of the time scales and persistency of tax evasion behavior

Figure 17.9 shows that within a 30-year time frame, most of our taxpayers that have initiated in tax evasion behavior are not caught and penalized for evasion. Amongst those taxpayers that were caught and penalized within the past six years, most remain compliant in the short-term. However, a significant proportion of those that are caught and penalized for tax evasion either do not become fully compliant immediately after a year that they were penalized or relapse back into tax evasion behavior within a couple of decades. The persistence of tax compliance behavior and tax evasion behavior of our taxpayers is shown by the histogram in Figure 17.10. We can deduce from this histogram that taxpayers that are persistent in being completely compliant always fully report 100% of their hideable income. Our model does not produce taxpayers that persistently (within a 30-year time frame) report more than 80% but less than 100% of their hideable income. Thus, these taxpayers can be considered "pathologically honest". Instead, taxpayers that can be considered persistent in their underreporting behavior do not necessarily always completely underreport all of their hideable income. Therefore, depending on how strict we define...
the persistency of tax compliance behavior, the sum of the two behaviors can range between just below 40% to just below 70%.

As a validation test, these results can be compared to what has been found using experiments methods. For example, Alm et al. found that taxpayers tend to exhibit "habit persistence", and that over multiple rounds participates in their experiment stabilized to an average filing rate between 58% and 68% \[7\]. These experiments tracked tax compliance persistence in terms of tax filing behavior. Since our model considers tax compliance behavior as how much is reported to the IRS and not whether or not a taxpayer files his/her taxes, this is not exactly the same tax compliance behavior. Thus, a direct comparison is not accurate. Nevertheless, our model shows significant "habit persistence" behavior which in general terms is consistent with the findings of Alm et al. \[7\].

### 17.4.5 Verification and validation of the distribution of reporting compliance rates

Figure [17.11] shows a histogram of the reporting behavior of our taxpayers found by the model using one of our calibrated best cases runs. As mentioned previously, our model reproduces the weighted empirical U-shaped distribution of reporting compliance rates found by Alm et al. \[6\]. This empirical distribution in shown Figure [17.4]. In this case, using the full model with heterogeneous taxpayers, the agreement between these two U-shaped distributions is stronger. As explained late in Section 20 on model calibration, this is not necessarily surprising as one of our calibration targets was to reproduce the empirical U-shaped distribution of reporting compliance as closely as possible. Nevertheless, as a model validation, we consistently observed that all of our model cases runs produced a U-shaped distribution of reporting compliance rates that overall
compared well with the ones reported by Alm et al. [6]. In these histograms, taxpayers that
under-report all of their hideable income are usually the majority, ranging from just below 50% to
80%. Between 9% to 20% report all of their hideable income. The remaining taxpayers are usually
spread uniformly in the middle. However, as shown in the example case run in Figure 17.11, we
often have very few taxpayers persistently reporting half of the hideable income. Most taxpayers
get polarized towards either end of the U-shaped distribution even if they do not completely
engage in either behavior. Figure 17.11 also illustrates that a significant majority of the taxpayer
that are caught for tax evasion within the past 6 years become compliant. However, this is a
short-term effect, and over time they slowly relapse back towards partial or complete tax evasion
behavior.

Figure 17.11 shows that most taxpayers in our model are non-compliant and this may suggest
a very large tax gap. However, the histogram provides the compliance level on the proportion
of income that can be hidden from the IRS. A large proportion of the income for most taxpayers in
our model is not hideable and is assumed to be reported automatically to the IRS. Figure 17.12 shows
a set of three histograms that show the compliance behavior of the taxpayers by their income,
hideable income and whether or not they are self-employed. Figure 17.13 shows the compliance
level of the taxpayers by the proportion of their total income, rather than by the proportion of
hideable income. These plots help us verify that taxpayers are generally compliant with regard to
their total income. The model produces a tax gap of about 16%. This value may appear to be small
but is instead relatively large since the maximum possible tax gap that our model can produce is
26.6%. This maximum tax gap would occur when all taxpayers underreport all of their hideable
income.
Figure 17.12: Histograms showing distribution of the percentage of hideable income reported by our taxpayers and by (a) the taxpayers’ income, (b) the taxpayers’ selfemployment status, (c) the taxpayers’ hideable income and (d) by the taxpayers’ number of alters they talk about taxes.

Figure 17.13: Histogram showing a distribution of the percentage of total income reported by our taxpayers once the model reaches a state of dynamic equilibrium. Taxpayers that have been never been caught for tax evasion within the past 30 years are shown in blue. Those that have been caught for tax evasion within the past 30 but not within the past 6 years are shown in green, and those caught within the past 6 years are shown in red.

17.4.6 Verification of the distribution of perceived levels of deterrence

Figure 17.14 shows the histogram of the perceived audit rate for our taxpayers and once the system has reached a stationary dynamic equilibrium. We immediately notice that the values of the
perceived audit rates are strikingly large. This is not surprising because the starting baseline perceived audit rates were informed directly from our ALP survey and were found to be very large. However, the model factors in these very large perceived audit rates and compensates their effect on compliance behavior using the model parameters $m_x$ and $s_x$. The histogram shows that the distribution of the perceived audit rate for taxpayers that have recently been audited (red) is more uniformly distributed, and has a higher mean value than that found for the distribution describing taxpayers that have not been audited in the past 30 years (blue). Hence, as expected by our model structure and design, taxpayers that are audited on average have larger perceived audit rates. Over time, their perceived audit rate relaxes back to their baseline value as shown by the distribution describing taxpayers that have been audited in the past 30 years but not in the past 6 years (green). Figure 17.15 shows the histograms of the perceived penalty rate for our taxpayers and once the system has reached a dynamic equilibrium. The distribution for the baseline perceived penalty rates is also informed from our ALP survey. In our model taxpayers that have in recent years (i.e., in the last 6 years) been caught for tax evasion and have been penalized (red) learn about the true penalty rate. Over time, their perceived penalty rate relaxes back to their baseline value as shown by the distribution describing taxpayers that have been penalized in the past 30 years but not in the past 6 years (green). However, in contrast, the audit rate, taxpayers learn about the true penalty rate after they get penalized and thus in our model, the timescales describing how the penalty rate relaxes back to the baseline value is much longer. We see this effect in Figure 17.15 as the form of the green histogram, more closely resembles the red histogram than the blue histogram, and this is in contrast to the corresponding forms of the histograms shown in Figure 17.14 which consider a shorter timescale for the perceived audit rate.

### 17.4.7 Verification and validation of the social network effects on tax compliance behavior

Using experimental methods, Alm et al has shown that taxpayers reporting decisions can strongly be affected by the reporting behaviors and the decision of people in their social network. Our model includes network mechanisms whereby reporting decisions and risk perceptions of taxpayers are influenced by the alters in their social network with which they talk about taxes. However,
Figure 17.15: Histograms showing the distribution of the perceived penalty rate expressed as a percentage and in base 10 logarithmic scale.

although the network mechanisms are present in the model they can be made to play a minor or insignificant role depending on the model parameter values that control them. Here we verify that our model does produce network effects and we illustrate and quantify its strength using our calibrated case run. As a validation test, we observe that network effects play a significant role in our model dynamics and is consistent with the findings by Alm et al.

Figure 17.16 shows a histogram for the values of $\Delta_N^{(i)}$ over all taxpayers. As defined in equation 6.1, the value for $\Delta_N^{(i)}$ for taxpayer $i$ is calculated as $\langle \Delta_P^{(j)} \rangle_{j \in J^{(i)}}$ which represents the average of the values of $\Delta_P^{(j)}$ over all his/her $J^{(i)}$ alters. This histogram shows that the majority of taxpayers that report nearly all of their hideable income (blue) have alters that have a very high compliance propensity. Similarly, the histogram shows that the majority of taxpayers that report nearly none of their hideable income (blue) have alters that have a very low compliance propensity. Taxpayers that report between 2 and 98% of their hideable income have on average a set of alters that generally have a moderate to low compliance propensity. This histogram suggests that there is a positive correlation between a taxpayer’s compliance behavior and the average compliance behavior of his/her alter. However, this correlation may not necessarily be caused by the model dynamics and instead could come from the assortative way taxpayers with similar $c_1^{(i)}$ threshold values are connected to the network. Thus, taxpayers with similar attitudes towards compliance may be connected to the network prior to any additional compounding network effects caused by the dynamics that would then lead to even higher levels of assortative mixing. Whether the observed network relationships are mainly caused by the initial network setup, leading to embedded pre-existing correlations or whether the dynamic network mechanisms are causing these correlations to emerge will be revisited in Sections 19 and 20 on model sensitivity and calibration. However, some early indication is provided by testing the correlations between the taxpayer’s dynamic attributes $\Delta_N^{(i)}$ and $\Delta_P^{(i)}$. In our case run, we found that this correlation was 73%, and this can be visually confirmed by heat map plot shown in Figure 17.17.

In our model, the weighted sum of $\Delta_P^{(i)}$ and $\Delta_N^{(i)}$ together with $\Delta_M^{(i)}$ strongly affect whether a compliant taxpayer will initiate in tax evasion behavior. To a lesser extent, this weighted sum

\begin{equation}
\Delta_N^{(i)} = \langle \Delta_P^{(j)} \rangle_{j \in J^{(i)}}
\end{equation}
Figure 17.16: Histogram showing the distribution of $\Delta_N^{(i)}$. The value of $\Delta_N^{(i)}$ is related to average compliance propensity over all the alters belonging to a given taxpayer $i$. The histogram distinguishes between taxpayers that report nearly all of their hideable income (blue), taxpayers that report nearly none of their hideable income (red), and taxpayers that report between 2 and 98% of their hideable income (green).

Figure 17.17: Heat map illustrating the correlation between taxpayers $\Delta_N^{(i)}$ and their $\Delta_P^{(i)}$. Based on the quintiles of the distribution of $\Delta_N^{(i)}$, we categorize the values $\Delta_N^{(i)}$ for each taxpayer in one of six categorical ranges. These categories are shown on the x-axis. We then did the same for $\Delta_P^{(i)}$ and these categories are shown on the y-axis. We then produced a matrix where each entry provided the number of taxpayers within each different combination of $\Delta_N^{(i)}$ and $\Delta_P^{(i)}$ category. The colors shown in the heat map are related to a relative value which we call "overlap". These are found by normalizing the values for each entry in our matrix by the entry with the maximum value.
determines the proportion of a taxpayer’s hideable income that s/he ends up reporting. As explained in our model description, this is because taxpayers that have initiated in tax evasion and have not been caught for their evasion will, over time tend to increasingly underreport larger proportions of their hideable income irrespective of their $\Delta_P^{(i)}$ and $\Delta_N^{(i)}$ values.

Figure 17.18: Heat map illustrating the correlation between taxpayers the average proportion of hideable income reported by a taxpayer’s alters and his/her proportion of reported hideable income. The overlap value was calculated using the same approach described in the caption for Figure 17.17.

Hence, the correlation between the average proportion of hideable income reported by a taxpayer’s alters and his/her proportion of reported hideable income is lower. This is confirmed by our case run which produced a correlation of 37%, and this correlation is illustrated by heat map plot shown in Figure 17.18. Notice, that the x-axis in this heat map is the average proportion of hideable income reported by a taxpayer’s set of alters. Thus, distribution of the proportions of hideable income reported by the alters of a taxpayer may largely differ from the average value. Therefore, we are also interested in the number of network edges connecting taxpayers with a subset of his/her alters that happen to have a similar proportion of hideable income reported. Hence, we can find the mixing matrix between the proportions of hideable income reported. Figure 17.19 shows a heat map illustrating the mixing matrix.

The level of assortative mixing between the proportion of hideable income reported by the taxpayers can be measured by network assortativity measure. One such measure, introduced by Newman [133] is found from the Pearson correlation coefficient of degree between pairs of linked nodes. Being a correlation coefficient, this value can range in [-1,1] from completed disassortativity mixing to complete assortativity mixing. For our model case run, we found that this measure gives a very low level of assortative mixing of 7%. However, since correlation is a measure of linear dependency it does not perform well when data is non-linear and highly skewed as shown by our histogram in Figure 17.16. One solution would be to calculate the Spearman correlation using the rank-transformed data. An alternative is to measure the deviation of the level of assortative mixing from what we would expect if the network connected taxpayers randomly according to an Erdős-Renyi graph. This measure is explained in appendix H. We find that the level of assortative
17.4.8 Verification of the how our model scales the outputs to the US national level

Section 9 describes how we can scale up our model outputs describing the dynamics of the aggregated quantities to be representative of US national levels for the year 2016. Here we use the same model case run that has reaches a state of dynamic equilibrium and scales the outputs up to the US national levels. Since our case run belongs to the set of calibrated cases, it agrees with the observed starting US tax gap. Figure 17.20 show scaled-up and representative of US national levels macro-level model dynamics.

Figure 17.21 shows the dynamics of the voluntary compliance rate that is related to the tax gap, and the marginal deficit with respect to the 2016 equilibrium value. Since the models reached a stage of dynamic equilibrium the plots illustrate the level of fluctuations the model produces for both the voluntary compliance rate and the marginal deficit. These fluctuations are high and are due to the fact that these model results consider our PN1 network with just above 1000 taxpayers. Hence, stochastic variability for such a low number of taxpayers in our model is relatively high and this is reflected when we scale up our numbers to be nationally representative. Nevertheless, we see that the voluntary compliance rate only fluctuates within a couple of percentage point and the fluctuations of the marginal yearly deficit with respect to 2016 are at most $50 billion which is roughly a tenth of the 2016 budget deficit of nearly $500 billion. Figure 17.22 shows a similar plot where instead of the marginal deficit it focuses on the debt that is attributable to tax compliance, and tracks this quantity with respect to the initial value for 2016.
Figure 17.20: Marco-level model dynamics outputs generated by one of our calibrated case runs and scaled up to be representative of US national levels. The macro-level quantities tracked in this plot include the total government expenses (which is constant by default), the total government revenues, the income tax revenues and the yearly budget deficit. Since the model has reached a stage of dynamic equilibrium and parameters and policy levers remain unaltered, the dynamics of these macro-level quantities remain stable.

Figure 17.21: Dynamics of the voluntary compliance rate (blue) and the marginal deficit with respect to the 2016 equilibrium value (red), and generated by one of our calibrated case runs and scaled up to be representative of US national levels.

Figure 17.23 shows the dynamics of the total audit costs and the dynamics of the unpaid taxes recovered by audit activities. Since the audit rates and the audit strategy is stationary over time, the audit costs from year to year do not change much. However, since the audit rate used in the model only 1% the model produces large stochastic variability in the amount of unpaid taxes recovered from the audited taxpayers. This large variability is due to a combination of using our
Figure 17.22: Dynamics of the voluntary compliance rate (blue) and the changes in debt (red) that is attributable to tax compliance with respect to the 2016 equilibrium value, and generated by one of our calibrated case runs and scaled up to be representative of US national levels.

PN1 network with just above 1000 taxpayers and to the fact, a low audit rate inherently produces a large uncertainty in the amount recovered by unpaid taxes from year to year.

Figure 17.23: Dynamics of the total audit costs (red) and the dynamics of the unpaid taxes recovered by audit activities (blue). We see that the recovered total tax revenues due to audit activities found by this case run of our model range between $10 and 30 billion a year which is between 2 to 6 times larger than the total costs of the audit activities. This indicates that the current audit rates and audit strategies are cost effective and there is room for a considerable increase in the audit rate.

Figure 17.24 shows the dynamics of the voluntary compliance rate and compares it to the dynamics of the ratio between the amount of recovered taxes and penalties due to audit activities and the costs of the audit. As expected, the voluntary compliance rate increases in the subsequent
years that follow a year when the IRS recovers large amounts of unpaid taxes and collects penalties through its audit activities. The plot also shows that when we consider the amounts generated by the applied penalties for tax evasion in addition to the recovered unpaid taxes, then these together are between 4 and 8 times larger than the total costs of the audit activities.

Figure 17.24: Dynamics of voluntary compliance rate (blue) and of the ratio between the amount of recovered taxes and penalties due to audit activities and the costs of the audit (red).

Figure 17.25 shows the dynamics of the voluntary compliance rate and compares it to the dynamics of the transformed debt to GDP ratio generated using the assumptions described in section 9.7 and equation 9.4. We refer to this debt as "transformed" since, unlike the debt calculations we have consider elsewhere in the model, this debt includes compound interest. Using this the simplistic assumption that GDP growth and interest rates on the debt remain unaltered at the 2016 levels, our case run shows that the debt to GDP can potentially reach levels as high as 1.5 within a century.

17.5 Verification with the larger PN10 network

In our model verification, validation and calibration, and later for the policy experiments, we mainly used solely the PN1 network, which considers a population of just over 1000 taxpayers. We could have used the PN10 and PN460 networks which consider a much larger number of taxpayers. However, running the model on these larger networks takes a significantly longer time. Our preference was to use the PN1 network so that run times were short, allowing us to consider many more independent case runs of our model, each with different settings. However, by considering the network with the smallest number of taxpayers the model produces trajectories of aggregated outcomes (e.g., the tax gap) with large chance-driven (i.e., stochastic) fluctuations. Here we compare the results of the model dynamics obtained using the PN10 network with just over 10,000 taxpayers to the model dynamics of the PN1 network using the same settings and other model specifications. We compare these dynamics for a set of 50 cases. The selection of these cases are specified in Section 20 on model calibration. The purpose of this verification is to build confidence in the model and in the results produced by using the PN1 network. However, we did not run our model using the PN460 network as the model runs even for 50 cases is very slow. We
leave this for future work where we will also explore more in depth how the stochastic fluctuations produced by the model dynamics scale with a larger network and number of taxpayers.

To make our comparison we ran each case for both the PN1 and PN10 network models to the stationary state. Once the dynamics reached the stationary state we tracked and recorded the dynamics for an additional 40 years. For example, for each case, we tracked the dynamics of the tax gap for both the PN1 and PN10 network models and obtained its time-averaged value and its standard deviation. We could then verify whether the time average values for the PN1 and PN10 were the same and how much smaller the variability of the trajectories of these aggregated outcomes were for the PN10 case compared to the PN1 case.

Table 17.1 shows the statistics of the differences across the 50 case runs of the time averages of five aggregated level outputs. These outputs include the tax gap and mean perceived audit rate in the population measured as percentages, and three assortativity network measures described in Section 17.4.7. The table shows that the aggregated dynamics using the PN10 network reproduces the PN1 network well. For example, the mean difference in tax gap is only 0.26% and varies between -0.88% and 1.78%. Other outputs also support this conclusion. Table 17.2 shows the standard deviation of the trajectories of the aggregated-level outputs over the 40-year dynamic for the PN1 and PN10 networks. As expected the variability in the PN10 network is lower than for the PN1 network.
Table 17.2: Table showing the standard deviations of the trajectories of the aggregated-level outputs over the 40-year dynamic for the PN1 and PN10 networks.

<table>
<thead>
<tr>
<th></th>
<th>tax.gap</th>
<th>mean.per.audit.rate</th>
<th>nn.over.nc</th>
<th>cc.over.cn</th>
<th>assortativity</th>
</tr>
</thead>
<tbody>
<tr>
<td>PN1</td>
<td>0.00609</td>
<td>0.00822</td>
<td>0.02550</td>
<td>0.07840</td>
<td>0.00943</td>
</tr>
<tr>
<td>PN10</td>
<td>0.00267</td>
<td>0.00243</td>
<td>0.00883</td>
<td>0.02360</td>
<td>0.00326</td>
</tr>
</tbody>
</table>

Figure 17.26: Comparison of the histograms of the proportion of taxable income reported for three example cases and using the PN1 network (left column) and the PN10 network (right column).

As an additional verification we compared the PN10 to the PN1 final year distribution of compliance rates for the 50 cases. Figure 17.26 show these distributions for 3 of the 50 cases. The comparison shows that the PN10 network does indeed reproduce the compliance levels observed in the PN1 network.

18 Model Outputs for Analyses and Calibration

As illustrated in the previous section, the model produces many different aggregated-level outputs and individual-level distributions of compliance behaviors and risk perceptions. In addition,
it produces scaled up aggregated outputs that can be interpreted at the US national-level, such as the deficit and the debt. However, for our uncertainty and sensitivity analyses, and our model calibration we focused on a few key model outputs which we can compare with known targets described in the literature. Table 18.1 lists and describes these model outputs and provides their target values and the sources.

<table>
<thead>
<tr>
<th>Name/Symbol</th>
<th>Description</th>
<th>Target</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Xi_t^{\text{[gross]}}$</td>
<td>Tax Gap given by the ratio between the net underreported tax and the total true tax liability.</td>
<td>17.1%</td>
<td>see section 9.2</td>
</tr>
<tr>
<td>$U_{\text{distance}}$</td>
<td>A distance measure between the histogram of the proportion of hideable income reported produced by the model and its target empirical histogram.</td>
<td>Figure 17.4</td>
<td>[6]</td>
</tr>
<tr>
<td>$H_{\text{persistance}}$</td>
<td>Percentage of taxpayers that do not change their compliance behavior and are persistent in the proportion of hideable income they report.</td>
<td>60%</td>
<td>[7]</td>
</tr>
<tr>
<td>$R_0$</td>
<td>Average percentage of hideable income reported by taxpayers prior to being caught and penalized for tax evasion.</td>
<td>28.6%</td>
<td>[20, 153]</td>
</tr>
<tr>
<td>$R_3$</td>
<td>Average percentage of hideable income reported by taxpayers three years after being caught and penalized for tax evasion.</td>
<td>63%</td>
<td>[20, 153]</td>
</tr>
<tr>
<td>$R_{\text{Annoyed}}$</td>
<td>Average percentage of hideable income reported by taxpayers three years after being audited but not found to be under-reporting on their taxes. We label these taxpayers as &quot;annoyed&quot; by being audited by the IRS.</td>
<td>35%</td>
<td>[20, 153]</td>
</tr>
<tr>
<td>$q_R$</td>
<td>The perceived audit rate of the taxpayers $\tilde{\eta}_t^{(i)}$ in year $t$ and how it relates to the initial mean perceived audit rate $\tilde{q}_0^{(i)}$ in the model. The ratio $q_R$ is defined as $\langle \tilde{\eta}_t^{(i)} / \tilde{q}_0^{(i)} \rangle$.</td>
<td>1.15</td>
<td>Estimated</td>
</tr>
</tbody>
</table>

Table 18.1: Model outputs and their target values used for the uncertainty and sensitivity analyses and model calibration.

The first output is the gross tax gap. Section 9.2 describes this output and provides how we estimated its target value from IRS reports. The second output is the distribution of the proportion of hideable income reported. Our target distribution is the one provided by Alm et al. [6]. The difference between the distribution produced by the model output and our target is summarized by a single number that summarizes the differences between the two distributions. Initially, we considered three standard measures, the Kolmogorov-Smirnov distance, the Kullback-Leibler and
the Jensen-Shannon divergences. However, after experimenting with these three measures we found that a simpler distance measure based on the sum of the squared error between the two histograms was best suited for our purpose. Specifically, this measure is found by first taking the difference in the counts expressed as a percentage of taxpayers belonging to the respective bins of the two histograms, and then squaring each of these differences and summing them. However, since our model consistently produces U-shaped histograms for taxpayer’s reporting compliance rates, and for added sensitivity we decided that the sum of the squared error should focus only on specific bins of the model histogram that are most sensitive and vary the most. These include the last two bins characterized by taxpayers with high compliance rates. The maximum value of the sum of the squared error for two bin measure is $2 \cdot 100^2$ and hence this measure ranges from $[0, 20,000]$.

The third model output listed in table 18.1 is the percentage of taxpayers that are persistent in their tax compliance behavior, and over a period of 30 years consistently report either 99% or more or 1% or less of their hideable income. Section 17.4.4 describes this measure in more detail and illustrates how this persistency measure may vary depending on how strict we define it. The estimated target for this measure is 60%. As mentioned in Section 17.4.4, this estimate is based on the findings reported in Alm et al.

The targets reported for the fourth, fifth and sixth model outputs are derived from the findings of the 2016 National Taxpayer Advocate (NTA) Audit Impact Study report Sebastian Beer and Ernard that has been summarized by Beer et al. We have described these targets previously in Section 12.3. Our model tracks $R_0$ which is calculated as the long-term (i.e., the entire duration of the model run) time average of the percentage of hideable income reported by taxpayers in the year they were audited and caught for tax evasion. Likewise, the model tracks $R_3$ which is calculated as the long-term average of the percentage of hideable income reported by the same set of taxpayers three years after being audited and penalized for tax evasion. The model output $R_{\text{Annoyed}}$ considers taxpayers that were audited but found not to be under-reporting on their taxes and calculates the long-term average of the percentage of hideable income they report three years after being audited. Note that in the year they get audited by the IRS, these taxpayers are not necessarily fully compliant. It is not unlikely that they are already under-reporting some of their hideable income but due to the IRS audit efficiency, none of it is detected.

Finally, the seventh model output computes the average overall taxpayers of the ratio between their present perceived audit rate and the initial perceived audit rate. This measure is expressed as $q_R = \langle \tilde{q}_i(i) / \tilde{q}_0(i) \rangle$. In our model, the initial/baseline perceived audit rate assumes that all taxpayers (and thus also all of their alters with which they talk about taxes) have never been audited in

---

67These three measures range from $[0, 1]$. The Kolmogorov-Smirnov statistic quantifies the maximum distance between the cumulative distribution function produced by the model and the cumulative distribution function of the reference distribution. The Kullback-Leibler and the Jensen-Shannon both provide related but slightly different measures of divergence (i.e., not a distance) between the entropy-like information content contained by the two distributions. These statistics are summarized in Veyrat-Charvillon and Staendaert and.

68Our model consistently produced U-shaped histograms of the taxpayer’s reporting compliance rates where (i) more than 45% of taxpayers reported none of their hideable income, and (ii) less than 20% of taxpayers reported all of their hideable income. However, the latter bin was found to be more sensitive and more important to get correct. For example, the targets for these two bins are 47.5% and 13.6% and a model producing 55% and 5% for these two bins represents a worse fit than a model producing 60% and 11%, even if the former has a lower sum of the squared error.

69We also considered differences that included four bins, which in addition to the last two bins included the middle two bins characterized by taxpayers that report between 40% and 60%.

70This is our most strict definition of persistence that tends to give the lowest possible estimate. Other persistence measures we considered included taxpayers that consistently report either 95% or more or 5% or less of their hideable income.
the past. As described in section 7.5.2, the taxpayers perceived audit rates are then modified by audit experiences during the model dynamics. As taxpayers experience IRS their perceived audit rate increases compared to the baseline. The output \( q_R \) measures the average increase in the perceived audit rate with respect to their baseline value over all our taxpayers. The difference \( q_R - 1 \) measure the increase in this proportion and is always positive. As described in Section 16.1, the initial/baseline perceived audit rates of our taxpayers in the model was informed based on ALP survey data and using the data of respondents that (i) never experienced an IRS audit, and (ii) do they know of taxpayers in the social network having ever being audited. However, as described in section 15.1, analysis of our ALP survey did not reveal statistical differences between the perceived audit rate of respondents who have had experiences with being audited or know of others who have had those experiences compared to respondents that never experienced an IRS audit. Thus, the ALP data would suggest that \( q_R \sim 1 \). However, it is reasonable to assume that the middle and long-term perceived audit rate of those that experienced an audit in the past do have an increased perceived risk of being audited. Short-term perceived audit rates may differ due to bomb-crater effects. Based on expert opinion and on looking at the variability in the ALP data of respondent’s perceived audit rates with and without audit experiences, we estimate that \( \langle \tilde{q}_R \rangle \) is about 15% greater than \( \tilde{q}_0 \). Hence, the target we specify for \( q_R \) is 1.15. However, due to the large uncertainty in this target, we do not put much weight on it.

19 Uncertainty and Sensitivity Analyses

Using the ranges of the model parameter and their specific sampling distributions that were specified in section Section 12.1 and in tables 12.4, 12.5 and 12.6 we used a Latin Hypercube Sampling (LHS) process to create a set of 2,500 cases each with a unique set of combinations of input parameter values. The systematic variation of the values of our parameters produced by the LHS process efficiently samples the entire input space within the ranges and sampling distributions specified for each model input parameter. Using these inputs for each case, we ran our model until it reached a stationary state of dynamic equilibrium. To test for stationary, the model checked whether the forty-year moving average of the first and second order changes in tax gap is within a specified and very stringent tolerance range. Appendix I describes the details of our test to check for stationarity. We ran 11 independent realizations for each case run. Each of these 11 realizations used the same set of input model parameters and initial conditions but used a different random number seed. Stochastic variability over each of these realizations produced slightly different model trajectory. Outputs for a given model case run were produced by averaging over these 11 realizations.

19.1 Uncertainty Analysis

Table 19.1 shows the distribution of our model outputs across our case runs and due to the uncertainty ranges described in our LHS of input parameter values. We see that most target output value is shown in table 18.1 fall within the ranges produced by our model outputs. For example, our target tax gap of 17.1% falls within the bounds of 4% and nearly 20% produced by our model.

---

71 For each case we ran one run starting from an initial condition with all taxpayers being compliant until it reached a stationarity state characterized by a dynamic equilibrium. We then used this final state as our new initial condition for other 10 runs, each using a different random seed. We then ran the original and the additional 10 model realizations forward using the same set of new initial conditions each producing slightly different trajectories due to changes in the chance effects.
Table 19.1: Statistics of the distribution of our model outputs across our 2,500 mode case runs.

<table>
<thead>
<tr>
<th>output</th>
<th>Min.</th>
<th>1st Qu.</th>
<th>Median</th>
<th>3rd Qu.</th>
<th>Max.</th>
<th>Mean</th>
<th>sd</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ξ [gross]</td>
<td>4.1</td>
<td>12.0</td>
<td>13.8</td>
<td>15.5</td>
<td>19.8</td>
<td>13.7</td>
<td>2.6</td>
<td>2500</td>
</tr>
<tr>
<td>U_distance</td>
<td>0.5</td>
<td>33.2</td>
<td>64.7</td>
<td>107.5</td>
<td>3889.0</td>
<td>199.1</td>
<td>489.5</td>
<td>2500</td>
</tr>
<tr>
<td>H_persistance</td>
<td>21.0</td>
<td>29.0</td>
<td>32.7</td>
<td>35.7</td>
<td>81.4</td>
<td>34.8</td>
<td>9.4</td>
<td>2500</td>
</tr>
<tr>
<td>R₀</td>
<td>19.8</td>
<td>29.9</td>
<td>34.4</td>
<td>40.1</td>
<td>54.5</td>
<td>35.0</td>
<td>6.7</td>
<td>2500</td>
</tr>
<tr>
<td>R₃</td>
<td>57.6</td>
<td>78.4</td>
<td>82.4</td>
<td>86.7</td>
<td>99.7</td>
<td>82.8</td>
<td>6.3</td>
<td>2500</td>
</tr>
<tr>
<td>R_Annoyed</td>
<td>0.0</td>
<td>3.7</td>
<td>7.4</td>
<td>11.9</td>
<td>29.4</td>
<td>8.0</td>
<td>5.5</td>
<td>2500</td>
</tr>
<tr>
<td>qᵣ</td>
<td>1.0</td>
<td>1.3</td>
<td>1.4</td>
<td>1.5</td>
<td>2.2</td>
<td>1.4</td>
<td>0.2</td>
<td>2500</td>
</tr>
</tbody>
</table>

However, for the case of $R_{\text{Annoyed}}$, its target value is above the maximum value that our model can achieve. Taxpayers that get audited in our model on average report less than 35% of their hideable income three years after being audited. The main reasons that our model cannot reproduce this target are it is calculated over all taxpayers independently of whether they were compliant or not. Thus, due to less than perfect IRS audit efficiencies in detecting unreported income assumed in our model, the noncompliant taxpayers that are audited and not caught for tax evasion, or that are caught but penalized for only a small part of the amount evaded, will continue to evade in future years at similar levels. If $R_{\text{Annoyed}}$ were calculated using a sample of taxpayers that are normally compliant or report most of their hideable income, then this value would be higher. Thus, the reason between the differences in the $R_{\text{Annoyed}}$ achieved by the model and our target value may be due to different samples or types of taxpayers considered and their starting compliance behavior. For the case of the histogram of the proportion of hideable income reported, the distance measure $U_{\text{distance}}$ ranges between 0.5 and just under 4,000. Our target distance is of course 0. However, considering that the maximum error is 20,000, we see that the error between our model distribution of hideable income reported generated by our model and the empirical distribution is relatively small.

### 19.2 Sensitivity Analysis

Table 19.1 only shows the ranges of model outputs produced using the set of model input values and the uncertainty in their ranges. However, we further want to know how the model outputs respond to variation in the input parameter values, and rank these inputs based on their leverage on the outputs. We do this via a sensitivity analysis. Our approach relies on using Classification And Regression Trees (CART) and Random Forest (RF) methods described in more detail in Appendix J. We also report Spearman’s partial rank correlation coefficients (PRCC) between the model outputs and the model inputs. However, in our sensitivity analysis,

---

72 The leverage of an input variable on an output refers to the measure of how sensitive the output values are on the input variable using our methods.

73 The partial correlation between a model output and a model input measures the degree of association between the two while controlling for the variability and the effects of the other model inputs. The partial correlation coefficient removes issues between confounding inputs. Therefore, it is generally a better measure than simple correlation and can reveal an association between an input and an output that is the inverse of that suggested by a simple correlation. For example, consider the following hypothetical scenario where taxpayers compliance behavior decreases as they earn more income. As a consequence, the IRS audits the higher income taxpayers more frequently. So higher income taxpayers (and those with low compliance rates) get audited more frequently than lower income taxpayers. The simple correlation between being non-compliant and being audited would be positive. In contrast, the partial correlation coefficient would show that for a given income level we would find that those who get audited more frequently become more compliant.
we were also interested in estimating how the model outputs respond to the overall tax rate, or rather any changes in the tax rate. We thus considered an additional model input to our LHS which measures changes in marginal tax rates. This input ranged uniformly between -8% to 8%. Thus, the marginal tax rates that are shown in table 8.1 were each reduced or increased by the same sampled percentage point from this input parameter. For example, if we sampled -8% the top marginal tax rate was reduced from 39.6% to 31.6%. Likewise, we reduced all the other marginal tax rates. All other input parameters and their sampling ranges were unmodified.

Figure 19.1: CART analysis of the tax gap model output.

Figure 19.1 illustrates an example CART diagram using our tax gap model output. It shows that relative to the tax gap, model cases can be optimally split into two groups or branches using the input parameter that measures the change in marginal tax rate and whether this input value is above or below -4.6%. The next most important parameter for the tax gap output is the overall audit rate. For the right-side branch, we see that this second-level split depends on whether the audit rate is above or below 0.9%. The resulting branches can then be further split based on the parameter $\beta_P$ which measure the relative importance of personal evaluations compared to network and media evaluations. For the right-most branch, we see that this third-level split depends on whether $\beta_P \leq 73\%$. We see that this resulting subset of cases with $\beta_P \leq 73\%$ groups 4% of all the cases we considered and the average tax gap over this subset is 13%. If instead $\beta_P < 73\%$ the split continues to a fourth-level and depends on the audit rate value again. For the right-most branch, we see that this fourth-level split depends on whether audit rate is smaller than 0.7%. We see that this resulting subset of cases with audit rates below 0.7% groups 7% of all the cases we considered and the average tax gap over this subset is 17%. This CART analysis illustrates that the change in tax rate, the audit rate and the value of $\beta_P$ are among the top drivers of the tax gap. This sensitivity is also confirmed by the 3D scatter plot shown in figure 19.2 and the parallel coordinate plot shown in figure 19.3.

---

74Since $\beta_P + \beta_N + \beta_M = 1$, the dependence on $\beta_P$ is really on how taxpayers weigh the different feedback components, and when media activity is low (the majority of times), it measures the relative importance of network and personal effects with respect to each other.
Figure 19.2: Scatter plot showing the distribution of our case runs according to the three input parameters that are among the top drivers of the tax gap, indicated by the color scale. These inputs include the change in the tax rate, the audit rate and the value of $\beta_P$. The plot provides a visual sense of the strength of the relationship between these three inputs and the tax gap output.

The sensitivity analysis shown in figures 19.1 - 19.3 was carried out for all of our model outputs. However, for our analysis, we relied more heavily on RF analysis rather than on CART. This is because although RF is a similar method to CART, it has desirable accuracies and robustness advantages over CART that are described in Appendix II. CART analyses are however more intuitive, and easier to interpret and to visualize. Both these methods produce overall measures of the leverage of each model parameter on the analyzed output. However, the magnitude of these values is hard to interpret as they rely on measures relating to node purity of the decision tree. Therefore, in our analyses, the sensitivity measures they produced were normalized to provide a relative score with respect to the input value that produced the highest sensitivity.

Figure 19.4 shows a sensitivity heat-map based on these relative scores. Here, the rows list seven main model outputs and the columns lists inputs with the most leverage on these outputs. For anyone single model output, this relative score can be interpreted across all model inputs. However, they cannot be compared across model outputs. To help us compare the sensitivity
Figure 19.3: Parallel coordinate plot illustrating the sensitivity of the tax gap output (shown on the left-hand side) on six input parameters. These six input parameters were chosen as they have the most leverage on the tax gap output as found from our sensitivity analysis. The plot illustrates the correlations between the tax gap output and each of the inputs, as well as the correlations amongst the inputs relative to the tax gap output. The values at the top and bottom of the output measure and each of the six inputs provides the range of values.

across model outputs we use the PRCCs. Our plot shows that overall the top input parameters and their ranges of values that have the most leverage on our model outputs include the changes in the tax rate, the audit rate $q_t$, the effects from the gamblers fallacy $\gamma_{GF}$, the weight $\beta_P$, the taxpayers compliance response to being audited and penalized $\vartheta$ (i.e., $v.PP$), and the morale half-life $\tau$.

Variations in the tax rate have a strong leverage on the tax gap, the histogram of the hideable income reported and the $R_0$ value. The weight $\beta_P$ (and hence the other weights $\beta_N$ and $\beta_M$), has a strong leverage on most of the model outcomes. Interestingly, it has a strong effect on $R_{\text{Annoyed}}$, the average percentage of hideable income reported three years after being audited and not being caught for tax evasion. The audit rate has a relatively strong leverage on the tax gap but counter-intuitively, our RF analysis suggests that it does not strongly affect the mean perceived audit rate in our taxpayers. The reason the RF analysis does not suggest a strong dependence may be due to the fact that the mode value considered for the audit rate is small and range of values we explored around the mode was relatively narrow. Consequently, the audit rate input value is not frequently used to split the data for the mean perceived audit rate output in the RF decision tree analysis. In contrast, the PRCC analysis does suggest a significant positive partial correlation between the audit rate and the mean perceived audit rate. Thus, although the audit rate is not frequently used to split the data it does, however, have a strong partial correlation with the mean perceived audit rate. Interestingly, both the RF and PRCC sensitivity analyses suggest that the mean perceived audit rate depends more strongly on the weight $\beta_P$, and hence on the weight $\beta_N$ and the effects of social network interactions. This is most likely due to the fact that the input parameter value for $\beta_P$ (and hence $\beta_N$) we used had a much large variability that the variability considered around the audit rate.

Our model conceptualized a gambler’s fallacy effect on the taxpayers, and we used the ALP to parameterize this effect. Importantly, this effect has yet to be observed in the real system with real
taxpayers. However, interestingly our sensitivity analysis shows that the mean perceived audit rate is most strongly affected by the gambler’s fallacy effect. Based on how the model was conceptualized and developed, this tells us that taxpayers who only recently initiated in tax evasion behavior largely increase their perceived audit rate (i.e., their fear of being audited) which may prolong their behavior of reporting most of their hideable income. The analysis also suggests that despite this strong effect on their perceived audit rate and on potentially prolonging partial compliance behavior, the gambler’s fallacy effect does not have much leverage on the tax gap, nor on the distribution of the proportion of hideable income reported. This indicates that aggregated tax revenues collected during the short period of increased fear of audits and of partial compliance of taxpayers who have recently initiated in tax evasion are not sufficiently large to play a major role on the tax gap and population-level compliance. Nevertheless, the strong leverage of this effect on the mean perceived audit rate suggests that it should be better studied in the real system as it could play a strong role on individual-level compliance behavior.

As expected the deterrence response parameter values $\theta$ (i.e., $v.PP$) which modulates and controls the effects that penalties have on taxpayers that are caught for tax evasion, has its strongest effect on the average proportion of hideable income reported by penalized taxpayers three years after being caught by the IRS (i.e., $R_3$). The PRCC analysis also suggests that it has a moderate

![Heatmap showing the sensitivity of the model outputs (rows) to six input parameters (columns) that have the most leverage on the outputs using a Random Forest method.](image_url)

**Figure 19.4:** Heatmap showing the sensitivity of the model outputs (rows) to six input parameters (columns) that have the most leverage on the outputs using a Random Forest method. Cells colored in dark red indicate the input that has the most leverage on a given output. Remaining cells are colored with a palette ranging from dark orange to light yellow and indicate decreasing leverage on the output measure relative to input with the most leverage. The numbers marked in black provide the numerical value of the relative leverage measure. The numbers marked in blue provide the PRCC value between the inputs and the outputs. The asterisk beside the PRCC values indicates a p-value scale lower than 1%.
to high leverage of \( \vartheta \) on the tax gap. The parameter \( \vartheta \) affects how long taxpayers that are caught and penalized for evasion remain compliant before they relapse back into tax evasion behavior. Therefore, it is expected to affect the individual-level compliance behavior. However, our analysis suggests that it’s aggregated effect is strong enough to affect the tax gap and population-level compliance.

Our analysis also suggests that the morale half-life \( \tau \), which controls how quickly taxpayers discount past experience and evaluations with respect to the present ones, has a strong leverage on the model dynamics. In particular, it affects the average proportion of hideable income reported by noncompliant taxpayers in the year that they get caught for tax evasion (\( R_0 \)). This indicates that it affects the average proportion of hideable income reported by noncompliant taxpayers and not just those who are caught by the IRS, and hence it strongly affects the tax gap.

The leverage of the other model parameters is summarized in Figures 19.5 and 19.6 which show results from our RF and PRCC sensitivity analysis using similar heat-maps. The first columns and inputs shown in Figure 19.5 are the same as those in Figure 19.4. The last column shows the aggregated effect of most of the remaining parameters put together, and it is used to get a sense of their collective leverage on the model outputs relative to the single parameters with the greatest leverages. These parameters with individually weaker leverages are shown in Figure 19.6 and are compared relative to each other. The analysis illustrated by these figures show that the deterrence response parameter \( m_x \) (and hence also \( s_x \)), the weight placed on the effect of receiving a tax return and the average number of alters of a taxpayer with which they talk about taxes have a relatively small PRCC values with the model outputs indicating weak correlations. The param-

Figure 19.5: Heatmap showing the leverages of four input parameters on our model outputs and previously shown in Figure 19.4. The last column shows the collective leverage of most of the remaining parameters with weaker leverages.
Figure 19.6: Heatmap showing the sensitivity of model outputs to the set of input parameters that have weaker leverages on the model outputs. The collective leverage of these input parameters relative to the inputs with greatest leverage is shown in Figure 19.5.

Figure 19.6: Heatmap showing the sensitivity of model outputs to the set of input parameters that have weaker leverages on the model outputs. The collective leverage of these input parameters relative to the inputs with greatest leverage is shown in Figure 19.5.

The approach we took to calibrate our model was to select the top 50 case runs out of our 2,500 set of cases runs that minimize an objective function represented by the weighted sum of the squared error between our model outputs and our calibration targets. Table 20.1 shows our main outputs together with the calibration target values first shown in Table 18.1. It also shows our choice for the calibration weights used to minimize our weighted sum of the squared error. The choice for the ordering of which calibration targets mattered the most was made by subjective judgments and expert consultation. The choice of the actual weights was made based on a trial
Table 20.1: The first three columns of this table list our seven main model outputs, our target values for the calibration and the respective weight used to minimize the weighted sum of the squared error between our model outputs and our calibration targets. The last three columns show results of a principal component (PC) analysis on the outputs of our 2,500 case runs illustrated in shown in Figure 20.5. They show the PC scale value that is used to normalize each output so that they can be compared on the same scale, the standard deviations of the PCs (i.e., the square roots of the eigenvalues of the correlation matrix), and the PC weight which expresses the variance of the PC as a percentage of the total variance.

and error approach. Our most important priority was to reproduce both the target tax gap and the empirical U-shaped distribution. However, at an early stage of our calibration exploration, we noticed that these two targets are not fully concordant and struggled against each other. This was later confirmed by a principal component analysis of our runs which is described later in this section. We also found that with the exception of $q_R$, the other five calibration targets often acted as tipping point parameters pushing the calibration to satisfy either one of our top two calibration targets. These five calibration targets do not have the same strength of evidence as the tax gap and the U-shaped distribution. They were either derived from interpretations of the results of published work (e.g., $R_0$, $R_3$ and $R_{\text{Annoyed}}$), taken from a not fully accurate interpretation of the published work (e.g., $H_{\text{persistance}}$) or based on expert opinion and weak indications from the ALP data (e.g., $q_R$). Therefore, we chose to provide relatively little weight to these five calibration targets.

### 20.1 Results in Matching our Calibration Targets

Figure 20.1 shows the ranking of our 2,500 cases in terms of the objective function and illustrates the range of values of our objective function for our top 50 cases with respect to the full 2,500 cases. These top 50 cases represent our set of calibrated cases. Table 20.2 shows the summary statistics

<table>
<thead>
<tr>
<th>output</th>
<th>target value</th>
<th>calibration weight</th>
<th>PC weight</th>
<th>scale</th>
<th>sdev</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Xi^{\text{[gross]}}$</td>
<td>17.10</td>
<td>20</td>
<td>51</td>
<td>2.60</td>
<td>1.90</td>
</tr>
<tr>
<td>$U_{\text{distance}}$</td>
<td>0.00</td>
<td>4</td>
<td>23</td>
<td>489.51</td>
<td>1.26</td>
</tr>
<tr>
<td>$H_{\text{persistance}}$</td>
<td>60.00</td>
<td>1</td>
<td>1</td>
<td>5.49</td>
<td>0.32</td>
</tr>
<tr>
<td>$R_0$</td>
<td>28.57</td>
<td>1</td>
<td>7</td>
<td>6.74</td>
<td>0.72</td>
</tr>
<tr>
<td>$R_3$</td>
<td>63.00</td>
<td>1</td>
<td>5</td>
<td>6.30</td>
<td>0.59</td>
</tr>
<tr>
<td>$R_{\text{Annoyed}}$</td>
<td>35.00</td>
<td>1</td>
<td>12</td>
<td>9.37</td>
<td>0.92</td>
</tr>
<tr>
<td>$q_R$</td>
<td>1.15</td>
<td>1</td>
<td>0</td>
<td>0.18</td>
<td>0.17</td>
</tr>
</tbody>
</table>

Table 20.2: Statistics of the distribution of our model outputs across our top 50 mode case runs.
of our top 50 calibrated cases which can be compared to the summary statistics for the full 2,500 cases shown in Table 19.1. These summary statistics and distributions for the top six outputs for both the full 2,500 cases and our calibrated cases are illustrated in the overlaid violin and box plots shown in Figure 20.2. We notice that despite the relatively strong weight place on the tax gap, our calibration process can only produce tax gaps that are generally smaller than our target of 17.1%. The error in reproducing the U-shaped distribution on the proportion of income reported is however strongly reduced. Interestingly, the mean value and range for $R_{\text{Annoyed}}$ is strongly reduced compared to the full 2,500 cases and is further away from our target value.

Figure 20.3 illustrates the correlation matrix of our model outputs produced by the full 2,500 cases. We conducted a principal component (PC) analysis of our outputs for both the full 2,500 cases and the reduced top 50 cases. Figure 20.4 shows a scatter plot of all the 2,500 cases relative to the two most important PCs. The plot also shows the set of out top 50 cases. From Figures 20.3 and 20.4, we immediately see in the negative correlation and unalignment of the tax gap and the U-shaped distribution explaining the struggle we mentioned between these two targets. Interestingly, we see a strong correlation and alignment between the U-shaped distribution and the persistence target. This is interesting as it suggests concordance between these two targets which is not surprising as they both come from results found by the same research group. We also notice concordance between $R_0$ and $R_3$ which also comes from the same study. Interestingly, these are not aligned with $R_{\text{Annoyed}}$ even though it comes from the same study. $R_{\text{Annoyed}}$ seems to be better matched with the tax gap target. This may be due to problems in the assumptions we used in calculating $R_{\text{Annoyed}}$. Results of this PC analysis are summarized in Table 20.1. The PC scores provide an indication of the importance of each output on the transformed axes and can be interpreted as a weight. Interestingly, the ranking of these weights is in good agreement with the calibration weights.

Figure 20.5 show the results of a PC analysis using just our top 50 cases. It illustrates how in

---

75 Notice that in Figure 20.2, the plot in the first row and second column shows $U_{\text{distance}}$ in a log base 10 scale. Our target for $U_{\text{distance}}$ is 0 which on a log scale is represented by $-\infty$, hence we indicate this by the dotted blue line at -1 (the lowest value is shown in this scaled plot).

76 Principal Component analysis is a form of multidimensional scaling. It linearly transforms the output variables into a lower dimensional space which retain the maximal amount of information about the variables.
Figure 20.2: Overlaid violin and box plots showing the form of the distribution, the quantiles and the mean (green dot) of six model outputs across our 2,500 case runs (red) and our top 50 cases representing our set of calibrated cases (blue). The horizontal dark blue lines represent our calibration targets. The weight placed on the calibration target is shown by the thickness level of the horizontal line.

Figure 20.3: Correlogram showing the correlation matrix of our model outputs found using our 2,500 case runs. Pie charts and values shown in blue/red indicate a positive/negative correlation. The pie charts and the transparency levels indicate the strength of the correlation. The diagonal terms label our model outputs and provide the range of values generated by our model runs.
Figure 20.4: Biplot illustrating the 2,500 case runs shown by the red dots in our scaled space using the first two PCs as our axes which respectively account for 51.33% and 22.5% of the variance in model outputs. Scale values used for this plot are shown in Table 20.1. The inner scatter plot shown using blue dots represent our set of 50 top cases.

Figure 20.5: Biplot illustrating the top 50 case runs shown by the blue dots in our scaled space using the first two PCs as our axes which respectively account for 44.52% and 27.43% of the variance in model outputs of our calibrated case runs. The plot is based on a different PC analysis using just the 50 top cases. Hence the PC axes/orientation and the positioning of the case runs change compared to Figure 20.4.
these top 50 cases the tax gap and the U-shaped distribution output align better due to the act of minimizing the objective function.

20.2 Analysis of the correlations in our model input parameters due to Calibration

Each of our 2,500 cases used a unique combination of model parameters. Some input parameters such as $\beta_N$ and $\beta_M$ were derived from the sampled $\beta_P$ due to the constraint that the three parameters sum to one. Thus, for the very few parameters that were derived from other input parameters we have a perfect correlation. Most input parameters were instead sampled independently from each other according to the LHS approach and therefore are not correlated. The process of calibration and selection of our top 50 case runs from the 2,500 cases introduces some correlations between the input parameters that arise due to the selection process. In this section, we present some results showing which correlations arise from this process.

Figure 20.6 illustrates the overlaid violin and box plots of 12 input parameters that have strong leverages on the model outputs. We see that the distribution for most input parameters does not change by much. The input parameters that are most affected by the calibration include $m_x$ (i.e., $m.qP$), $\beta_P$ and $\tau$ (i.e. morale.half.life).

We computed the correlation matrix between our input parameter values sampled from the LHS for all our 2,500 and verified that the off-diagonal terms showed little or negligible correlations values. We then computed the correlation matrix for our inputs parameter values belonging to the 50 case runs representing our calibrated cases. By comparing the two correlation matrices we identified which parameters had the biggest and most significant changes in correlation values and selected this subset of the parameters. Table 20.3 shows the initial correlation matrix found for this subset of input model parameters for our full 2,500 cases. Table 20.4 shows the final correlation matrix found for the same subset of input model parameters for the top 50 cases.

![Figure 20.6: Overlaid violin and box plots showing the form of the distribution, the quantiles and the mean (green dot) of twelve model inputs across our 2,500 mode case runs (red) and our top 50 cases representing our set of calibrated cases (blue).](image-url)
Table 20.3: Correlation matrix for a subset of input parameters using the full 2,500 cases runs.

<table>
<thead>
<tr>
<th></th>
<th>detection.eff</th>
<th>c2</th>
<th>m.qP</th>
<th>beta.personal</th>
<th>rate.refund.movement</th>
</tr>
</thead>
<tbody>
<tr>
<td>detection.eff</td>
<td>1.00</td>
<td>-0.06</td>
<td>0.13</td>
<td>0.14</td>
<td>0.11</td>
</tr>
<tr>
<td>c2</td>
<td>-0.06</td>
<td>1.00</td>
<td>-0.14</td>
<td>-0.20</td>
<td>-0.12</td>
</tr>
<tr>
<td>m.qP</td>
<td>0.13</td>
<td>-0.14</td>
<td>1.00</td>
<td>0.25</td>
<td>0.17</td>
</tr>
<tr>
<td>beta.personal</td>
<td>0.14</td>
<td>-0.20</td>
<td>0.25</td>
<td>1.00</td>
<td>0.21</td>
</tr>
<tr>
<td>rate.refund.movement</td>
<td>0.11</td>
<td>-0.12</td>
<td>0.17</td>
<td>0.21</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Table 20.4: Correlation matrix for a subset of input parameters using the 50 case runs representing our calibrated cases.

<table>
<thead>
<tr>
<th></th>
<th>detection.eff</th>
<th>c2</th>
<th>m.qP</th>
<th>beta.personal</th>
<th>rate.refund.movement</th>
</tr>
</thead>
<tbody>
<tr>
<td>detection.eff</td>
<td>1.00</td>
<td>-0.23</td>
<td>0.31</td>
<td>0.43</td>
<td>0.44</td>
</tr>
<tr>
<td>c2</td>
<td>-0.23</td>
<td>1.00</td>
<td>-0.32</td>
<td>-0.45</td>
<td>-0.15</td>
</tr>
<tr>
<td>m.qP</td>
<td>0.31</td>
<td>-0.32</td>
<td>1.00</td>
<td>0.70</td>
<td>0.45</td>
</tr>
<tr>
<td>beta.personal</td>
<td>0.43</td>
<td>-0.45</td>
<td>0.70</td>
<td>1.00</td>
<td>0.41</td>
</tr>
<tr>
<td>rate.refund.movement</td>
<td>0.44</td>
<td>-0.15</td>
<td>0.45</td>
<td>0.41</td>
<td>1.00</td>
</tr>
</tbody>
</table>

These correlations can be more easily visualized using a PC analysis which is shown in Figure 20.7. We find that after the calibration process the parameters $\beta_P$ and $m_x$ are positively correlated, and negatively correlated with $c_2$. We also find that the parameter describing the detection efficacy and the rate of refund return movement are correlated. It is very hard to explain why these correlations arise and interpret their meaning because these input parameters represent very different processes and enter the model in very different sections and modules of the model. They are interesting but require more analysis.

Figure 20.7: PC analysis of five input parameters that showed the biggest increase in correlation after the process of calibration.
Policy Experiments

21 Simple intervention and "what if" scenarios

We begin by considering a set of simple policy experiments and "what-if" scenarios. These experiments are simple because they explore changes in one of the following levers: the tax rate, the audit rate, the income distribution and the employment type distribution. Moreover, in each of these experiments, we use our calibrated set of cases and run the system to the new stationary state. Proceeding case by case, we compare outputs of interest obtained at the new stationary state to the starting condition. Therefore, in these first simple experiments, we do not focus on the dynamics and on how soon the new stationary state is reached. Rather, we run the system for as long as it is needed to reach the new stationary state where the system is again in a new state of dynamic equilibrium.

For the reasons described in Section 17.5, most of our policy experiments and "what if" scenarios described in this section make use of our PN1 synthetic population which considers a network of just 1195 taxpayers. Consequently, even when our model reaches a new stationary state, the year-to-year fluctuations due to the chance-driven (i.e., stochastic) effects are significant and cannot be neglected. To reduce inaccuracies that result from these stochastic fluctuations, we chose to compare the 40-year time averages of the model outputs in the initial and final stationary state. The initial 40-year average is taken during the dynamics describing the stationary state with the status-quo settings This is done before we implement the policy changes. The final 40-year average is taken after the system reaches its new stationary state using the new settings describing the policy or "what if" scenario.

21.1 "What if" scenario with more equitable income distribution

Here we consider a simple hypothetical "what if" scenario where we change the income distribution of our taxpayers and make it look like a more equitable income distribution of a different country. The income distribution of our taxpayers in our model has a Gini coefficient of 48%. Countries with a more equitable income distribution have a lower Gini coefficient. For example, Sweden’s Gini coefficient is 25%77. To transform the income distribution to one that is representative of Sweden, we first downloaded freely available income data for Sweden obtained by the International Social Survey Programme78. We found that the income ranges from $1,440 to $1,296,000 with a median of $77,790. However, our analysis showed that the Gini coefficient for this data is 35%. Thus, we made a further transformation of the income data in order to get a Gini coefficient which is closer to 25%. The transformation we used can be expressed as

\[ I_{\text{Sweden New}}^{(i)} = I_{\text{Sweden Old}}^{(i)} - 0.32 \cdot \left[ I_{\text{Sweden Old}}^{(i)} - \text{Median}(I_{\text{Sweden Old}}^{(i)}) \right] \tag{21.1} \]

Countries with large Gini coefficient have a higher level of inequitable income distributions. As quoted by Wikipedia at https://en.wikipedia.org/wiki/List_of_countries_by_income_equality the CIA’s 2014 estimate of the Gini coefficient for the US was 47%. This compares well with those quoted in the literature. This is not surprising as our income distribution was informed by census data. For Sweden, the same source lists a Gini coefficient of 25%.

77 The file ZA6770_v2-1-0.dta.zip containing the income data for Sweden from the International Social Survey Programme: Work Orientations IV - ISSP 2015 is available at https://dbk.gesis.org/dbksearch/sdesc2.asp?no=6770. The number of respondents in this data is 1,162. The income data is provided in Swedish Krona (SEK), which we converted to US dollars using an exchange rate where 1 SEK is equal to $0.12.
where $I^{(i)}_{\text{Sweden, Old}}$ is the income in dollars of the respondent in the data and $I^{(i)}_{\text{Sweden, New}}$ is the transformed income. This transformation has the effect of tightening the income distribution around the median. However, it strongly affects the income range especially on the lower side as after applying this transformation the income varies between $22,640 and $902,900. This transformation does however provide an income distribution with a Gini coefficient of 25%. In summary, our new income distribution is informed by two separate statistics regarding the income distribution for Sweden: (i) survey data on income and (ii) the Gini coefficient. Matching the two statistics produces an income distribution with desired features for our hypothetical "what if" experiment, but due to the strong effect of the transformation on lower incomes, we cannot claim that this is a very accurate representation of the income distribution for Sweden.

Using this new income distribution we made a second transformation in order to recenter the mean income distribution and make it align with our starting mean distribution for the US. This is because our aim was to produce an income distribution for the US which looks like the one for Sweden but without modifying the aggregated sum our taxpayer’s incomes. Finally, we assigned our taxpayers in our model a new income value taken from this new income distribution, making sure that the assignment preserved the original ordering of the taxpayer’s incomes. Using the new taxpayer’s incomes we calculated their new effective tax rate using the tax schedule shown in Table 8.1. These new incomes together with the new effective tax rates were used to run our hypothetical "what if" experiment. Figure 21.1 shows both our starting income distribution and our transformed income distribution for this experiment.

Table 21.1 summarizes the findings of this experiment. The model suggests that the with a more equitable income distribution, similar to that found in Sweden there would be just over $200 billion less income tax revenue and the average taxpayer would be paying between $1,278 and $1,452 less in taxes every year. This reduction is due to a decrease in the overall effective tax rate by roughly 2 percentage points (pp). Despite this decrease in the effective tax rate and the total income revenues collected, our model suggests that the tax gap would potentially increase. Overall our 50 case runs the median increase in the tax gap is of 0.78 pp. However, this could range from a decrease of 0.87 pp to an increase of as much as 2.31 pp. To get an insight of why this experiment produces an increase in the tax gap we refer back to Figure 21.1. The histograms show the income distributions before and after the experiment and provide the compliance level of hideable income reported for each income bin. In our starting income distribution, we have many
taxpayers with very low incomes. The compliance levels for these taxpayers are relatively low. Under the new scenario, the compliance level of these same set of taxpayers does not increase. This is because these low-income taxpayers remain low propriety to the IRS for an audit and consequently are not selected for an audit that frequently. However, since these taxpayers are now earning more income their contribution results in reducing the tax revenue and increasing the tax gap. On the other side of the income spectrum, taxpayers with high incomes remain the high priority for IRS audits. Their, compliance level also does not strongly change by this experiment. However, since they now earn less income their contribution also results in lower tax revenues.

The limitations of this experiment are that it only considers changes in the income distribution to reflect one which is more equitable, such as that found in Sweden. However, these hypothetical scenarios are not fully representative of the Swedish settings, because it is known that taxpayers in Sweden have a strong trust in the way their government uses taxpayers money to fund public services and social welfare programs. Arguably there is also a stronger trust that other taxpayers are paying their fair share in taxes. Consequently, a more representative experiment that transforms the US settings to those found in Sweden would need to consider changes in behavioral parameters and the distribution of taxpayer’s behavioral attributes. Most notable of these would include the distribution of the $c_1^{(i)}$ values and the $\beta_N$ parameter.

### 21.2 Changes in the proportion of self-employed taxpayers

We considered the hypothetical scenario where the proportion of self-employed taxpayers is significantly reduced. In our fifty calibrated cases, the self-employed taxpayers represent 16.8% of the total number of taxpayers. In our model, the self-employed attribute that specifies whether a taxpayer is self-employed depends on another attribute called the `self.employment.propensity`. This attribute was described in Section [15.2](#). We consider two hypothetical scenarios. The first reduces the number of self-employed taxpayers to 10% and the other to 5%. Our approach was to use the `self.employment.propensity` to randomly select self-employed taxpayers in our starting synthetic population and change their status from self-employed to non-self-employed (i.e., salaried). When the taxpayers were selected, we used Table [11.3](#) to change their proportion of hideable income from 100% to a new lower value based on their income and changed their `self-employed` attribute. We note that this process does not impose that the mixing matrix between self-employed and non-self-employed taxpayers is maintained at the level observed in our survey. Thus, as the number of taxpayers that are self-employed is reduced, the proportion of links connecting self-employed with other self-employed taxpayers also decreases.

Table [21.2](#) shows summary statistics of the change in the tax gap under a 10% and 5% self-employment scenarios. Surprisingly, we find that for both scenarios the tax gap increases com-

<table>
<thead>
<tr>
<th>Outputs</th>
<th>Min.</th>
<th>1st Qu.</th>
<th>Median</th>
<th>3rd Qu.</th>
<th>Max.</th>
<th>Mean</th>
<th>sd</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Change in Income Tax revenues (billions $)</td>
<td>-221</td>
<td>-213</td>
<td>-210</td>
<td>-208</td>
<td>-206</td>
<td>-211</td>
<td>3.61</td>
<td>50</td>
</tr>
<tr>
<td>Change amount of taxes paid per taxpayer per year ($)</td>
<td>-1452</td>
<td>-1321</td>
<td>-1276</td>
<td>-1235</td>
<td>-1105</td>
<td>-1278</td>
<td>70.69</td>
<td>50</td>
</tr>
<tr>
<td>Change in Tax Gap (%)</td>
<td>-0.87</td>
<td>0.27</td>
<td>0.78</td>
<td>1.13</td>
<td>2.31</td>
<td>0.75</td>
<td>0.70</td>
<td>50</td>
</tr>
</tbody>
</table>

Table 21.1: Changes in outputs of interest using a new income distribution with a Gini coefficient of 25% and informed using the income data of Sweden.
pared to the *status-quo* scenario. For the 10% scenario, the average increase across our 50 case runs is by 2.38 pp. For the 5% scenario, the average increase is smaller and equal to 1.23 pp. These

<table>
<thead>
<tr>
<th>Change in Tax Gap in percentage points</th>
<th>Min.</th>
<th>1st Qu.</th>
<th>Median</th>
<th>3rd Qu.</th>
<th>Max.</th>
<th>Mean</th>
<th>sd</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scenario 1 with 10% self-employed taxpayers</td>
<td>0.72</td>
<td>2.07</td>
<td>2.36</td>
<td>2.75</td>
<td>4.20</td>
<td>2.38</td>
<td>0.75</td>
<td>50</td>
</tr>
<tr>
<td>Scenario 2 with 5% self-employed taxpayers</td>
<td>-1.20</td>
<td>0.74</td>
<td>1.12</td>
<td>1.86</td>
<td>2.90</td>
<td>1.23</td>
<td>0.86</td>
<td>50</td>
</tr>
</tbody>
</table>

Table 21.2: Summary statistics of the change in the tax gap between the hypothetical scenarios and the *status-quo* scenario.

results are surprising because a decrease in the self-employment rate means that there are more salaried taxpayers that are characterized by significantly smaller hideable incomes. Hence, the set of taxpayers that were self-employed in the *status-quo* scenario and were in the position to under-report all of their income can no longer do so under these new scenarios, and this would lead us to think that we should get an increase in tax revenues. Tables 21.3 and 21.4 provide some indication of why this is not the case for the 10% and the 5% scenario respectively. These tables

<table>
<thead>
<tr>
<th>Averages</th>
<th>NSE</th>
<th>SE</th>
<th>SE→NSE</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>status-quo</em> income reported (%)</td>
<td>91.9</td>
<td>46.1</td>
<td>47.8</td>
</tr>
<tr>
<td>SE10 income reported (%)</td>
<td>89.7</td>
<td>54.8</td>
<td>89.7</td>
</tr>
<tr>
<td><em>status-quo</em> hideable income reported (%)</td>
<td>13.3</td>
<td>46.1</td>
<td>47.8</td>
</tr>
<tr>
<td>SE10 hideable income reported (%)</td>
<td>11.0</td>
<td>54.8</td>
<td>7.8</td>
</tr>
<tr>
<td><em>status-quo</em> amount of income reported (k$)</td>
<td>64.0</td>
<td>41.1</td>
<td>32.0</td>
</tr>
<tr>
<td>SE10 amount of income reported (k$)</td>
<td>61.1</td>
<td>48.0</td>
<td>51.7</td>
</tr>
</tbody>
</table>

Table 21.3: Average compliance levels of the taxpayers belonging to the three mutually exclusive groups described in the main text for the 10% self-employed scenario (labeled here as SE10).

<table>
<thead>
<tr>
<th>Averages</th>
<th>NSE</th>
<th>SE</th>
<th>SE→NSE</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>status-quo</em> income reported (%)</td>
<td>91.9</td>
<td>44.2</td>
<td>47.9</td>
</tr>
<tr>
<td>SE5 income reported (%)</td>
<td>89.7</td>
<td>58.8</td>
<td>89.5</td>
</tr>
<tr>
<td><em>status-quo</em> hideable income reported (%)</td>
<td>13.3</td>
<td>44.2</td>
<td>47.9</td>
</tr>
<tr>
<td>SE5 hideable income reported (%)</td>
<td>10.8</td>
<td>58.9</td>
<td>7.6</td>
</tr>
<tr>
<td><em>status-quo</em> amount of income reported (k$)</td>
<td>64.0</td>
<td>36.4</td>
<td>38.2</td>
</tr>
<tr>
<td>SE5 amount of income reported (k$)</td>
<td>61.6</td>
<td>41.2</td>
<td>60.0</td>
</tr>
</tbody>
</table>

Table 21.4: Average compliance levels of the taxpayers belonging to the three mutually exclusive groups described in the main text for the 5% self-employed scenario (labeled here as SE5).

consider three sets of mutually exclusive taxpayers: (1) taxpayers that are non-self employed in both scenarios (NSE), (2) taxpayers that are self employed in both scenarios (SE), and (3) taxpayers that are self employed in the *status-quo* but are non-self-employed in the new hypothetical scenario (SE→NSE). For each set of taxpayers and for each scenario the tables shows the (a) average proportion of their total income reported to the IRS, (b) the average proportion of their hideable income reported to the IRS, and (c) the average amount of income reported to the IRS in units of thousands of dollars. From Table 21.3 for the 10% scenario, we see that although the proportion
of hideable income reported by the SE→NSE taxpayers dramatically falls from 47.8% to just 7.8%, as expected the amount and the proportion of their total income reported to the IRS increases. A similar effect can be seen in Table 21.4 for the 5% scenario. This result can be explained by the way we model the IRS audit strategy which was described in Section 8.3. For each income bracket, the model assumes that the IRS preferentially audits those that are self-employed. Therefore, under the new scenario these taxpayers are audited by the IRS less frequently, and hence although these taxpayers increase their absolute level of compliance (i.e., based on the amount and proportion of income reported), they decrease their relative level of compliance (i.e., based on the proportion of hideable income reported).

Under these hypothetical scenarios, there are fewer taxpayers that are self-employed (i.e., the SE). Consequently, these taxpayers are targeted more frequently by the IRS which leads to increases in their compliance rate both in absolute and in relative terms. Salaried taxpayers that are non-self-employed in both scenarios (i.e., the NSE), have on average a marginal decrease in both their absolute and relative compliance levels. This is because a larger proportion of non-self-employed under the new scenarios (i.e., the NSE and the SE→NSE) leads to a decrease in the chance that any given non-self-employed taxpayer is selected for an audit. This, in turn, leads to a decrease in the individual-level compliance among these taxpayers. Although this decrease is marginal, the proportion of non-self-employed under the new scenarios is 90% and 95% respectively. Hence, at the aggregated population-level this is much larger.

The aggregated population-level effect can be seen in Figure 21.2. The histograms show the distribution of the taxpayers with respect to the change in the amount of income they report to the IRS with respect to the status quo scenario. The three different groups of taxpayers are shown in different colors. As expected, we see that the SE→NSE taxpayers report more of their income under the new scenarios. However, their aggregated effect (i.e., the sum of the red bars) is overwhelmed by the marginal decrease in what the NSE taxpayers report (i.e., the sum of the blue bars). Hence, this leads to an overall decrease in compliance level and an increase in the tax gap.

Comparison of the histograms shown in Figure 21.2 also help explain why the decrease in overall compliance is more pronounced in the 10% scenario than in the 5% scenario. In the 5%
scenario, there is a larger proportion of SE→NSE, and these are reporting more of their income. The compliance level of the NSE and SE do not change much compared to the 10% scenario and hence the overall effect is that the tax gap does not decrease by as much.

To verify our interpretation we ran two additional hypothetical scenarios. The first considered a 0% proportion of self-employed taxpayers and the second considered a 32% proportion of self-employed taxpayers. We found that for the 0% scenario the tax gap on average decreases marginally. For the 32% scenario the tax-gap increased and ranged between 20% and 28%.

### 21.3 Changes resulting from the 2018 Tax Reform

President Donald Trump signed the Republican tax bill into law on December 22nd, 2017. Starting from 2018, income earned will be taxed using different tax brackets with lower marginal tax rates. In the months preceding the tax reform, there was much deliberation on what would the new tax brackets and marginal rates.

#### 21.3.1 Initial proposed changes to the tax schedule

Table 21.5 shows an earlier version of what was proposed. The President was considering reducing the number of tax brackets to just three.

<table>
<thead>
<tr>
<th>Marginal Tax Rate (%)</th>
<th>Head of Household ($ k)</th>
<th>Married Filing Jointly ($ k)</th>
<th>Married Filing Separately ($ k)</th>
<th>Single ($ k)</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>[0, 37.5)</td>
<td>[0, 75)</td>
<td>[0, 37.5)</td>
<td>[0, 37.5)</td>
</tr>
<tr>
<td>25</td>
<td>[37.5, 112.5)</td>
<td>[75, 225)</td>
<td>[37.5, 112.5)</td>
<td>[37.5, 112.5)</td>
</tr>
<tr>
<td>33</td>
<td>[112.5, 1e+06)</td>
<td>[225, 1e+06)</td>
<td>[112.5, 1e+06)</td>
<td>[112.5, 1e+06)</td>
</tr>
</tbody>
</table>

Table 21.5: Initial proposal for the US Tax brackets and marginal tax rates for 2018.

Using the initially proposed changes to the tax schedule shown in Table 21.5, we ran each of our fifty calibrated cases to their new stationary state and computed the changes in the income tax revenues and the tax gap. Table 21.6 summarizes the findings of this experiment. Had the initially proposed changes in the tax schedule be voted into law, our model suggests that it would

<table>
<thead>
<tr>
<th>Outputs</th>
<th>Min.</th>
<th>1st Qu.</th>
<th>Median</th>
<th>3rd Qu.</th>
<th>Max.</th>
<th>Mean</th>
<th>sd</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Change in Income Tax revenues (billions $)</td>
<td>-109</td>
<td>-105</td>
<td>-104</td>
<td>-103</td>
<td>-102</td>
<td>-104</td>
<td>1.78</td>
<td>50</td>
</tr>
<tr>
<td>Change amount of taxes paid per taxpayer per year ($)</td>
<td>-697</td>
<td>-596</td>
<td>-575</td>
<td>-535</td>
<td>-374</td>
<td>-567</td>
<td>61.5</td>
<td>50</td>
</tr>
<tr>
<td>Change in Tax Gap (%)</td>
<td>-2.08</td>
<td>-0.55</td>
<td>-0.20</td>
<td>-0.03</td>
<td>0.96</td>
<td>-0.29</td>
<td>0.57</td>
<td>50</td>
</tr>
</tbody>
</table>

Table 21.6: Changes in outputs of interest using the initial proposal for the US Tax brackets and marginal tax rates for 2018.

have lead to a reduction of just over $100 billion of yearly income tax revenues. This represents a significant increase in the yearly deficit. However, our model does not account for the potential increase in productivity and GDP, and ultimately in increased taxable earnings that some believe will indirectly result from these reduced tax rates. Across our 50 cases runs, the results suggest that the average taxpayer would save between $374 to $697 in fewer income taxes every year. These results further suggest that although tax compliance is expected to increase, the average reduction of the tax gap across all our case runs is small, and on average it is only about 0.3 pp. However, the range across the 50 runs is large suggesting the decrease in tax gap could be as much as 2 pp, but could also produce no reduction at all or even an increase in tax gap of close to 1 percentage point.

21.3.2 The 2018 tax reform

However, the tax schedule is shown in Table 21.6 was not the final version that was approved. Instead, when the new law was signed in, it was decided to retain the seven income tax brackets but changes were made in the income ranges and marginal tax rates. Table 21.7 shows the new tax schedule that is now applied to all declared income earned by taxpayers starting from 2018.

<table>
<thead>
<tr>
<th>Marginal Tax Rate (%)</th>
<th>Head of Household ($k)</th>
<th>Married Filing Jointly ($k)</th>
<th>Married Filing Separately ($k)</th>
<th>Single ($k)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>[0, 13.6)</td>
<td>[0, 19.05)</td>
<td>[0, 9.526)</td>
<td>[0, 9.526)</td>
</tr>
<tr>
<td>12</td>
<td>[9.526, 51.8)</td>
<td>[19.05, 77.4)</td>
<td>[9.526, 38.7)</td>
<td>[9.526, 38.7)</td>
</tr>
<tr>
<td>22</td>
<td>[38.7, 82.5)</td>
<td>[77.4, 165)</td>
<td>[38.7, 82.5)</td>
<td>[38.7, 82.5)</td>
</tr>
<tr>
<td>24</td>
<td>[82.5, 157.5)</td>
<td>[165, 315)</td>
<td>[82.5, 157.5)</td>
<td>[82.5, 157.5)</td>
</tr>
<tr>
<td>32</td>
<td>[157.5, 200)</td>
<td>[315, 400)</td>
<td>[157.5, 200)</td>
<td>[157.5, 200)</td>
</tr>
<tr>
<td>35</td>
<td>[200, 500)</td>
<td>[400, 600)</td>
<td>[200, 300)</td>
<td>[200, 500)</td>
</tr>
<tr>
<td>37</td>
<td>[500, 1e+06)</td>
<td>[600, 1e+06)</td>
<td>[300, 1e+06)</td>
<td>[500, 1e+06)</td>
</tr>
</tbody>
</table>

Table 21.7: US Tax brackets and marginal tax rates for 2018.

<table>
<thead>
<tr>
<th>Outputs</th>
<th>Min.</th>
<th>1st Qu.</th>
<th>Median</th>
<th>3rd Qu.</th>
<th>Max.</th>
<th>Mean</th>
<th>sd</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Change in Income Tax revenues (billions $)</td>
<td>-235</td>
<td>-226</td>
<td>-223</td>
<td>-221</td>
<td>-218</td>
<td>-224</td>
<td>3.82</td>
<td>50</td>
</tr>
<tr>
<td>Change amount of taxes paid per taxpayer per year ($)</td>
<td>-1358</td>
<td>-1292</td>
<td>-1247</td>
<td>-1190</td>
<td>-1101</td>
<td>-1242</td>
<td>62.88</td>
<td>50</td>
</tr>
<tr>
<td>Change in Tax Gap (%)</td>
<td>-1.70</td>
<td>-0.89</td>
<td>-0.37</td>
<td>0.13</td>
<td>0.74</td>
<td>-0.40</td>
<td>0.61</td>
<td>50</td>
</tr>
</tbody>
</table>

Table 21.8: Changes in outputs of interest using the initial proposal for the US Tax brackets and marginal tax rates for 2018.

80Since our model does not account the tax deductions but only the effects of strictly applying the tax schedule, these savings could be greater.
We reran our model using this new tax schedule using our 50 calibrated cases as initial conditions. Results for this experiment are shown in Table 21.8. Our runs used the PN1 network which, as mentioned previously produces results that have a higher degree of stochastic variability. Therefore, in order to verify and confirm our results, we decided to repeat the experiment using the PN10 network. Results using the PN10 network are shown in Table 21.9. Our results

<table>
<thead>
<tr>
<th>Outputs</th>
<th>Min.</th>
<th>1st Qu.</th>
<th>Median</th>
<th>3rd Qu.</th>
<th>Max.</th>
<th>Mean</th>
<th>sd</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Change in Income Tax revenues (billions $)</td>
<td>-241</td>
<td>-229</td>
<td>-228</td>
<td>-225</td>
<td>-224</td>
<td>-228</td>
<td>3.81</td>
<td>50</td>
</tr>
<tr>
<td>Change amount of taxes paid per taxpayer per year ($)</td>
<td>-1330</td>
<td>-1296</td>
<td>-1279</td>
<td>-1257</td>
<td>-1171</td>
<td>-1275</td>
<td>28.58</td>
<td>50</td>
</tr>
<tr>
<td>Change in Tax Gap (%)</td>
<td>-1.28</td>
<td>-0.47</td>
<td>-0.27</td>
<td>-0.08</td>
<td>0.22</td>
<td>-0.29</td>
<td>0.28</td>
<td>50</td>
</tr>
</tbody>
</table>

Table 21.9: Changes in outputs of interest using the initial proposal for the US Tax brackets and marginal tax rates for 2018.

using the PN10 network are very similar to those obtained with the PN1 network. Ranges and means of the three outputs of interest are similar. However, we do see that the variability in the predicted reduction in tax gap is reduced. These model results suggest that the increase in the annual deficit due to the reduction in tax revenues could range between $228 to $241 billion. Again, this is assuming that the incomes of the taxpayers do not change due to these tax cuts. This reduction would lead to savings between $1,275 and $1,358 for the average tax payer. However, as before using the initial tax schedule, our model predicts that the reduction in tax gap will be modest and on average it will decrease by just 0.3 pp.

We compared these results to those found in our sensitivity analysis described in section 19.2. As illustrated in Figure 19.3 in our sensitivity analysis we varied the effective tax rate by ±8 percentage point. That is a very big variation in the tax rate. Our analysis found that the tax gap strongly responsible for changes in the effective tax rate when it is allowed to vary within this large range. In our model, the new tax schedule leads to an overall reduction in the effective tax rate of 2 pp. Therefore, although the reduction is taxed by the new law are large, they are relatively small compared to the range of tax cuts and increases we considered in the sensitivity analysis.

### 21.4 Cost effectiveness in direct benefits of increasing the audit rate

During model verification, we found that the recovered tax revenues due to IRS audits are much larger than the overall direct costs of audits. Figure 17.23 suggests that the current audit rates and audit strategies are cost effective and there is room for a considerable increase in the audit rate. Here we use our model to explore the cost-benefit of increasing the audit rate. Currently, the overall audit rate is less than 1%. In this experiment, we use a simple root-finding method to find the audit rate where the aggregated direct audit costs balance the aggregated recovered tax revenues (i.e., our root). This expresses a direct benefit of increasing the audit rate as we are only interested in the amount of recovered tax revenues and not in the additional indirect benefits of higher overall compliance levels.

---

81 This result can be compared against other predictions. For example, Greszler estimates that a single individual earning the median wage of $50,000 will take-home an additional $1,100 in income per year, and a married couple with $75,000 in income will take-home up to $2,000 in income per year. The original source can be found at [https://www.heritage.org/taxes/commentary/how-tax-reform-will-put-more-money-your-wallet-year](https://www.heritage.org/taxes/commentary/how-tax-reform-will-put-more-money-your-wallet-year).
We first specify the lower and upper values of the audit rate range. The lower bound is given by the status-quo audit rate and the upper bound audit rate was 9%. Each of our 50 calibrated cases is run to equilibrium using the new audit rate, and we take a 40-year average of the model outputs, including all the costs and recovered revenues at this new equilibrium. We then verify that using an audit rate of 9% produces average audit costs that are much larger than the average recovered tax revenues for all case runs. Then using a bisection method approach, we iteratively bisect our audit rate interval and select a subinterval in which the root must lie for further processing. Therefore, in our first iteration, we consider a 5% audit rate and run all of our 50 cases to a new equilibrium. For each case, we determine whether our root is bounded by the subinterval [1,5] or by the subinterval (5,9]. Depending on which subinterval is selected our second iteration considers either an audit rate of 3% or an audit rate of 7% for further model runs. This process is repeated until the error, expressed by the absolute value of the difference between the average audit costs and the average recovered tax revenues, is smaller than a chosen threshold. The threshold value is expressed as a percentage of the recovered tax revenues and we chose to set it at 5%. If the number of iterations used is large and the stopping criteria are not met, we terminate the iterative process and record the last iteration that produced the lowest error.

Table 21.10 shows the results of our findings. The first row in the table shows the summary statistics of the error in our stopping criteria for our root-finding method in matching average audit costs balance to average recovered tax revenues. Most cases reach our specified stopping criteria with an error of less than 5%. The second row shows summary statistics based on our outputs:

<table>
<thead>
<tr>
<th>Outputs</th>
<th>Min.</th>
<th>1st Qu.</th>
<th>Median</th>
<th>3rd Qu.</th>
<th>Max.</th>
<th>Mean</th>
<th>sd</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>error in matching (%)</td>
<td>0</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>7</td>
<td>3</td>
<td>2</td>
<td>50</td>
</tr>
<tr>
<td>audit rate increase (%)</td>
<td>148</td>
<td>172</td>
<td>194</td>
<td>239</td>
<td>544</td>
<td>228</td>
<td>93</td>
<td>50</td>
</tr>
<tr>
<td>audit cost increase (%)</td>
<td>123</td>
<td>151</td>
<td>163</td>
<td>200</td>
<td>452</td>
<td>196</td>
<td>77</td>
<td>50</td>
</tr>
</tbody>
</table>

Table 21.10: Summary statistics of our audit rate experiments that explore the direct benefits of increasing the audit rate.

root value and reports the percentage increase in the audit rate. On average, the audit rate can be increased by up to 200% before the costs of the audit rate begin to be greater than the recovered tax revenues. Therefore, this find suggests that the audit rate can be increased on average to about 3%. The third row shows summary statistics of the percentage increase of the audit rate costs.

We note that as the audit rate increases, compliance levels increase and this leads to diminishing marginal direct benefits in increasing the audit rate. Therefore, the increase in the audit rate is not as large as one might first anticipate.\(^{82}\) Table 21.11 provides summary statistics of the change in the audit costs. We see that the audit cost would increase on average from a total of $2.84 billion to $8.41 billion. However, this increase in the audit rate would produce indirect benefits by increasing the overall compliance level of the taxpayers. Table 21.11 shows that the average increase in income tax revenues due to this increase is $122.71 billion. Table 21.12 provides summary statistics of the indirect benefits by reporting the change in the tax gap, the change in indirect benefits by reporting the change in the tax gap, the change in

\(^{82}\)For example, looking at Figures 17.23 and 17.24, we may anticipate that the audit rate could be increased by a factor ranging between 4 and 8. However, as the audit rate increases the red line in Figure 17.23 moves up and at the same time, the blue line moves down. The two lines meet somewhere in the middle characterized by an average audit rate of 5%.
mean perceived audit rate and the change in the mean perceived penalty rate with respect to the status-quo. We find that the policy of increasing the audit rate to the point where the audit costs balance the recovered tax revenues would reduce the tax gap to a level ranging between 5 and 8%. This increased compliance is due to both the direct and indirect effects of deterrence. The increase in the indirect effects of deterrence can be measured by the change in the mean perceived audit and penalty rates in the population. We see that the mean perceived audit rate increases by on average 2.8 percentage points (pp). This is surprisingly low considering that the audit rate has increased by more than three fold. However, this expresses a change in the mean perceived audit rate in the population and we must bear in mind that the audit rate remains relatively low and most taxpayers are unlikely to experience an audit within a short to medium term time frame. Thus, for those taxpayers that do not experience an audit, their perceived audit rate only increases.
due to social network effects. The change in the perceived penalty rate is statistically negligible. This is because in our model an increase in the audit rate does not change the way the IRS selects taxpayers to be audited. Thus, the same set of taxpayers are more frequently audited. These taxpayers will generally be aware of the true penalty rate due to a previous experience with the IRS audit in their lifetime as a taxpayer. Hence, they will not change their perceived penalty rate by much.

21.5 Cost effectiveness in indirect benefits of increasing the audit rate

We follow from our previous analysis to further explore the cost-benefit of increasing the audit rate. In the previous section, we used a simple root-finding method to find the audit rate where the audit costs balance the direct benefits represented by the recovered tax revenues. As expected, at this new increased audit rate there are also indirect benefits represented by a high overall compliance rate of the taxpayers. This suggests that if we further increase the audit rate, the audit costs will on average be greater than the recovered tax revenues. However, increasing the audit rate may still be cost-effective and desirable if any additional increases in the audit costs provide larger amounts of tax revenues due to increased compliance. Here, we estimate the audit rate where the marginal increase in audit costs balance the marginal increase in income tax revenues due to the increase compliance.

Our approach does not use the same bisection root-finding method employed previously. Rather, we sequentially increase the audit rate in predefined steps of 1 pp. The reason we proceed in a sequential manner is that we need to calculate the marginal increase in costs and the marginal increase in income tax revenue at the new audit rate relative to the previous iteration that considered a slightly lower audit rate. As before, each case is run to the new equilibrium and audit costs and income tax revenues at this new equilibrium are computed by taking a 40-year average. The stopping criteria for our search are specified by the audit rate where the marginal increase in audit cost is greater than the marginal increase in income tax revenues. Presentation of our results follows the same format as that presented in the previous section.

Table 21.13 shows the results of our findings. The first row in the table shows the summary statistics of the error in our stopping criteria. This error is given by the absolute value of the difference between the marginal increase in income revenue and the marginal increase in audit cost, expressed as a percentage of the marginal increase in income revenue. We immediately see that this error can be quite large. The mean is 18% but for one of our 50 case runs the stopping criteria produce an error as big as 54%. The reason the error is not insignificant is that our sequential search proceeds using relatively coarse audit rate increments of 1 percentage point. More accurate estimates can be obtained using finer increments. The second row in Table 21.13 shows that

<table>
<thead>
<tr>
<th>Outputs</th>
<th>Min.</th>
<th>1st Qu.</th>
<th>Median</th>
<th>3rd Qu.</th>
<th>Max.</th>
<th>Mean</th>
<th>sd</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>error in matching (%)</td>
<td>0</td>
<td>7</td>
<td>17</td>
<td>27</td>
<td>54</td>
<td>18</td>
<td>14</td>
<td>50</td>
</tr>
<tr>
<td>audit rate increase (%)</td>
<td>714</td>
<td>1056</td>
<td>1208</td>
<td>1375</td>
<td>1576</td>
<td>1196</td>
<td>212</td>
<td>50</td>
</tr>
<tr>
<td>audit cost increase (%)</td>
<td>661</td>
<td>957</td>
<td>1074</td>
<td>1157</td>
<td>1470</td>
<td>1063</td>
<td>183</td>
<td>50</td>
</tr>
</tbody>
</table>

Table 21.13: Summary statistics of our audit rate experiments that explore the benefits of increasing the audit rate.
on average the audit rate can be increased by 12 pp to roughly 13% before the marginal increase audit cost is greater than the marginal increase in income tax revenues. The third row shows the increase in the audit costs.

Table 21.14 shows that the average increase in audit rate costs is $30 billion and this could range between $18 billion to $40.6 billion. The average increase in income tax revenues is $214 billion and ranges between $169 billion and $283 billion. Table 21.15 provides summary statistics of the change in the tax gap, the change in mean perceived audit rate and the change in the mean perceived penalty rate with respect to the status-quo. We find that the policy of increasing the audit rate to the point where the marginal increase in audit cost balances the marginal increase in income revenue would on average reduced the tax gap by 12.55 percentage point and the final tax gap would range between 0.17% and 3.02%. In Section 9.2 we mentioned that the component of the US
The gross tax gap for individual income taxes is $319 billion and this represents 17.1% of the expected individual income tax revenues. Therefore, it is legitimate to ask why Table 21.14 does not show slightly higher increases in income tax revenues closer to $300 billion when Table 21.15 shows that the tax gap is reduced on average to a very low value of just under 1%, effectively indicating near to full compliance. The reason can be found by looking at our first row of Table 21.15. As explained in Section 20, the tension between calibrating our model to a 17% tax gap and getting the right U-shaped distribution of the proportion of hideable income reported produced a set of 50 calibrated runs where the starting tax gap describing our status-quo cases ranges between 11.4% and 17.12% with an average of 13.4%. Hence, for most of our case runs the tax gap in the status-quo amounts to a lower than $319 billion.

Table 21.15 also shows the change in the perceived audit rate and the perceived penalty rate. We see that the mean perceived audit rate increases by on average 9.2 pp and ranges between 25.8% and 57.6%. We also see that on average there is a slight increase in the perceived penalty rate. For the same reason explained in the previous section, this could be statistically negligible. Previously, we argued that since an increase in the audit rate does not change the way the IRS selects taxpayers to be audited and the same set of taxpayers are selected for an audit more frequently. This implies that an increase in audit rate doesn’t necessarily imply an increase in the perceived penalty rate. However, when the audit rate is increased from under 1% to about 13%, then the validity of this assumption begins to break down since the IRS may not only increase the frequency that in audits taxpayers that were previously selected, but may also increase in the pool of taxpayers that it targets for an audit. This implies that more taxpayers get audited and thus learn about the true penalty rate. Further analysis of the model data can reveal and validate our explanation.

### 21.6 Intermittant changes in the audit rate

Generally, people find the uncertainty related to the possibility that something bad might happen to be more stressful than believing that it is definitely going to happen [47]. Increased uncertainty about the possibility of a bad event is generally more stressful. Consequently, to reduce this increased stress they might try harder to decrease or avoid uncertainty. There is a reason to believe that this applies also to IRS audits. One way of increasing the uncertainty and stress level in taxpayers is to change the audit rate from year to year with the hope that taxpayers would balance this increased uncertainty by increasing their compliance. This can be done in a way that the average audit rate and audit cost per year do not change. To test this hypothesis using our model we construct four different hypotheticals “what if” scenarios. The first three of these increase the audit rate by a factor of 2, 4 and 8 respectively, but also decrease the frequency that in any year the IRS audits taxpayers. Therefore, to keep the average audit rate unchanged, the IRS only carries out tax audits periodically every 2, 4 and 8 years. In the fourth hypothetical scenario, we assume that the period between successive audit years is randomly selected in the range [1,10] and the audit rate during the year that the IRS audits the taxpayers’ increases accordingly.

Table 21.16 shows the change in the tax gap associated with each of our four experiments. Results from our model experiments suggest that taxpayers do not change their average compliance behavior under increased uncertainty in the audit rate. These results should not be interpreted as evidence that increased uncertainty in the audit rate has little or no effect on compliance in the real world. Instead, they may reveal a limitation of the assumption and behavioral mechanisms considered by the model. These results are not surprising based on how the model was constructed. The time-average personal evaluation of the taxpayers in our model would not change due to the policy changes considered in this experiment. However, network and media effects could change.
Table 21.16: Summary statistics showing the change in the tax gap due to a policy of intermittent audit years with respect to the status-quo scenario.

The mechanism behind this is that knowing that many of your alters were audited in any one year could result in a super-additive effect compared to knowing that the same set of alters was audited in different years. Moreover, an increased audit rate in any one year could trigger a media feedback effect due to their monitoring of the moving average of the aggregated number of taxpayers that get audited and penalized. However, the time-averaged effects reported in Table 21.16 suggest that these effects and the way they have been parametrized in our model are not strong enough to increase overall compliance.

22 Combining fiscal and deterrence intervention policies

In the previous section, we explored a set of different hypothetical "what if" scenarios. In particular, we separately considered changes in fiscal policy and deterrence policy. Moreover, we compared the final new stationary state of the different "what if" scenarios to the starting status-quo condition. However, we did not consider the time taken to reach the new equilibrium and dynamics of the changes in compliance level and the changes in the debt. In this section, we explore policies that change both the tax rate as well as the audit rate and we consider the dynamics of the policy intervention. In particular, we use our model to address our primary research question described in Section 4. In the first part of the research question, we search for policies that can increase overall compliance and reach a specified target. Obviously, any increased compliance target can be reached if we are prepared to significantly decrease the tax rate and/or significantly increase the audit rate. However, these corner-solutions are extreme and could significantly decrease tax revenues and increase audit costs. These could lead to significantly larger yearly deficits and unmanageable increases in the debt. Instead, we are interested in socially acceptable and economically sustainable strategies that can reach our desired level of compliance within a given time frame. By socially acceptable strategy we mean policies that do not increase the audit rate to levels that can be perceived to be absurd and oppressing. By economically sustainable strategy, we mean a policy that does not decrease the tax revenues and increase budget deficits to the point that the debt becomes unsustainable. In the second part of our research question, we search for policies that can maintain the increased level of compliance obtained in the first part, and at the same time pay back the additional or marginal accrued debt that was developed during the first part. This research question tests the hypothesis that a combination of an increased audit rate and a
decreased tax rate leads to increased compliance in fewer years while minimizing the marginal debt compared to just an increased audit rate or a decreased tax rate on their own.

22.1 Part 1: Increasing the voluntary compliance rate

We start by considering the first part of the policy question and search for policies that can increase the voluntary compliance rate by using a combination of increasing the audit rate and decreasing the tax rate while minimizing the marginal debt. Using an experimental design, we explore which combinations work. Our chosen target is 95% voluntary compliance rate or more. Equivalently, policies that manage to decrease the tax gap to 5% or less satisfy these criteria.

Operationally, our experimental design varies both the tax rate and the audit rate in discrete intervals compared to the status-quo settings:

- The tax rate is varied between 75% and 100% in intervals of 5%. Hence, the tax rate varies by a multiplicative factor.

- The increase in the audit rate varied between 0 and 5 pp using seven uneven intervals. Thus the policy’s audit rate is the baseline audit rate in the status-quo, plus the increase in the audit rate. Hence, the audit rate varies by an additive factor.

Each experiment is initialized using our set of 50 calibrated case runs describing the status-quo settings. Therefore, the experiment considers a set of 48 experiments for each of our 50 cases. Using the new policy settings and initial conditions we run the system forward and record whether

![Diagram of voluntary compliance rate and marginal debt over years]

Figure 22.1: Example trajectory of the voluntary compliance (blue) and the debt (red) for part 1 of our policy experiment. In this case, the tax rate is reduced by a multiplicative factor of 85% and audit rate is increased by 5 pp. During our runs, we record the year when the voluntary compliance rate reaches 95% and the marginal debt. These are used as inputs to our part 2 policy runs. In this example, the year when the voluntary compliance rate reaches the compliance target is 2066 and the marginal debt in just under $5 trillion.
the compliance reaches our chosen target. If the target is reached we record what year it was reached for the first time, and what was the associated marginal accrued debt with respect to the status-quo. For the policy experiment to meet the search criteria and be considered a successful outcome, it is required to have reached the compliance target within a time window of 100 years. Moreover, a successful outcome should not include false positives. These occur when the target is reached due to a stochastic fluctuation but compliance stays below the target most of the time. We filter out the false positives by running the policy experiment to its new stationary state and checking whether the compliance trajectory is above the target for at least 50% of the time.

Figure 22.1 illustrates one example dynamics for the case where the tax rate is reduced by a multiplicative factor of 85% and audit rate is increased by 5 pp. We see that compliance under such an aggressive policy, voluntary compliance increases rapidly in the first decade of the policy. However, although compliance reaches very close to the target compliance level it only crosses it in the year 2066. The marginal debt in 2066 is close to $5 trillion. As can be seen in the figure, the dynamics are allowed to run well beyond the time it first reaches the compliance target. As explained, this is done to determine whether the voluntary compliance rate truly reached our
compliance target and to filter out false positives.

Figure 22.3 shows a heat map illustrating the outcome of our analysis of all the case runs of our experimental design. Each cell corresponds to a given policy setting in terms of changes in tax rate and audit rate. As mentioned previously, each policy experiment was run for all of our 50 cases. The color of a given cell provides an indication of the number of cases that reached the compliance target and that satisfied the additional criteria. Cells shown in red indicate that none of the 50 cases reached the compliance target. Cells shown in lighter colors ranging from orange to green indicate an increasing number of cases that reached the compliance target. For the cases that do reach the compliance target we report three average quantities in each cell:

1. The number of years needed to reach the compliance target.
2. The marginal debt in the year that the target was reached. Note, that negative numbers here indicate a marginal surplus.
3. The percentage of time the compliance level remains above the compliance target, which is used to filter out false positives.

The top two row in Figure 22.3 suggest that as long as the tax rate is not reduced by a multiplicative factor of less than 95% and the audit rate is increased by at least 2.75 pp, then on average

Figure 22.3: Heat map summarizing results of our analysis of the policy case runs. The color code provided in the legend gives the percentage of cases in each cell that reached the 95% compliance target. The numbers in the cells are explained in the main text. This analysis considered 2400 cases since each cell describing a policy setup considers 50 cases.
Figure 22.4: Heat map summarizing results of our analysis of the policy case runs. This heat map is similar to the one shown in Figure 22.3 but focuses only on cases that strictly combine both a change in the audit rate and a change in tax rate and that does not lead to marginal surpluses. Out of the 2400 cases considered in Figure 22.3 this plots shows 386 cases.

These policies would lead to a marginal surplus rather than a debt by the year the compliance target is reached. As the audit rate increases, the average number of years needed to reach the compliance target is reduced. Consequently, also the average surplus decreases since there were fewer years.

Figure 22.4 shows a similar heat map as Figure 22.3 that focuses only on cells that have cases that reach the compliance target and that strictly combine both a change in the audit rate and a change in tax rate and that does not lead to marginal surpluses. Therefore, average quantities may differ from those shown in Figure 22.3, especially for the top rows as cases that lead to surpluses are not used here in the calculation of the average quantities. The reason for showing this additional heat map is that case runs considered here are used as inputs to part 2 of our policy experiment where we explore policies that can maintain the achieved increased level or compliance while paying back the marginal debt.

Both Figures 22.3 and 22.4 suggest that if the tax rate is reduced by a multiplicative factor of 90% or less, then we can still reach our compliance target as long as the audit rate is increased by at least 2.5 pp. As expected, the lower the tax rate the greater the marginal debt. Surprisingly, the year that the compliance reaches its target for the first time is not very sensitive to the tax rate but it is very sensitive to the audit rate. Therefore, for a given audit rate (i.e., for any given column in the figure), reducing the tax rate by a multiplicative factor of less than 90% generally does not decrease the number of years needed to reach the compliance target but it does significantly increase the marginal debt. Instead, increasing the audit rate has generally a stronger effect in decreasing the
years needed to reach the compliance target and consequently on decreasing the debt.

We note that these conclusions are stronger for cells that contain more cases to average over on (i.e., green, light green and yellow cells). Cells which have 10% or less (i.e., 5 cases or less) are subject to increased variability due to fewer cases to average over. Another important point to notice is that our calibrated cases representing the status-quo do not all start from the same tax gap as our initial condition. Instead, as shown in Table 20.2 these range between 12.3% and 17.3%. Obviously, cases that start with a lower tax gap have a higher chance of reaching the compliance target of 95% and will generally reach it sooner producing lower marginal debt. Therefore, cells which have 10% of cases or fewer cases usually contain cases that start with lower tax gaps, closer to 12.3%. This is not an insignificant consideration and it is important to bear in mind in order to better interpret these results.

Despite these considerations and for the purpose of our policy goals, these results and analyses suggest that increasing the audit rate has a stronger effect than decreasing the tax rate and the reported averages lead us to believe that this is a robust finding. They further reject the hypothesis we stated at the start of this section and suggest that an increase in the audit rate on its own is sufficient at increasing compliance while also produce a marginal surplus. Any possible reduction in the number of years in reaching the compliance target by also decreasing the tax rate has a small effect, while it has a large effect in decreasing the small marginal surplus and instead can lead to large marginal deficits. As a consequence, exploration of part 2 of our research question becomes a partially superfluous exercise.

22.2 Part 2: Maintaining the higher compliance rate and reducing the marginal debt

Policies in part 2 of our research question aim to maintain the increased level of compliance achieved in part 1 while reducing the marginal debt back to $0. We saw in Figure 22.3 in the previous section that part 1 policies that increase the audit rate without changes in the tax rate can reach our desired level of increased compliance and produce a marginal surplus. An additional small reduction in the tax rate by at most 5% can help reach the compliance target in fewer years while still producing a marginal surplus. As described at the end of the previous section these policies are preferable than further reducing the tax rate which leads to marginal debts. Further, since they do not produce a marginal debt there is no part 2 policy component for these cases. However, for completeness, we will explore part 2 policies for all the other less desirable cases found in part 1. These produce a marginal debt and are illustrate in Figure 22.4.

Operationally, the experimental design for part 2 considers multiple scenarios for each successful outcome in part 1. The different plots that are shown in each row of Figures 22.5 and 22.6 illustrates an example trajectories of successful outcomes taken from part 1. Figure 22.5 shows shows the compliance trajectory and Figure 22.6 shows the related trajectory for the marginal debt. Two additional examples are shown in Figure 22.7. As can be seen from the trajectories of the compliance level, for each case we run twenty new policies starting from the year the compliance level reached our target value of 95%. These new trajectories are shown by different colors. As can be seen from the trajectories of the marginal debt, some policies are able to maintain the achieved level of increased compliance while slowly paying back the accrued marginal debt. Eventually, these policies can repay the accrued marginal debt. However, other policies do not decrease the marginal debt despite maintaining the increase compliance level. The latter set of policies are undesirable. Here, we are interested in exploring and identify the set of desirable policies that maintain the increased compliance and can decrease the marginal debt back to zero within a time frame of at most 100 years.

In order to pay back the marginal debt that was accrued during part 1, these policies need to
Figure 22.5: Compliance trajectories for an example case run. The dotted horizontal line shows the 95% compliance target. The dotted vertical line separates the our part 1 from our part 2 policy exploration. As explained in the main text, for each case explored in part 1 we have 20 case explored in part 2 (shown by different colored trajectories).

Figure 22.6: Marginal debt trajectories for an example case. The horizontal line in our marginal debt trajectory plots show the $0 debt level and helps identify desirable policies that can maintaining the increased compliance of 95% achieved in part 1 while paying back the accrued marginal debt in part 2.
Figure 22.7: Compliance (column 1) and marginal debt (column 2) trajectories for two additional example cases, each shown on different rows.

potentially increase the tax rate. At the same time, in order to prevent the compliance level to decrease below our achieved 95% compliance target, they potentially need to further increase the audit rate. The increase in the tax rate that we consider ranges between 100% and 115% of the tax rate used during our part 1 experiment using 5% intervals (i.e., a total of four values). For example, consider a part 1 policy experiment that reached our compliance target using a tax rate equal to 85% of the rate used by the status quo. If we consider a part 2 policy that uses a 110% increase in tax rate, the new tax rate of part 2 represents $1.10 \times 0.85$ that was used in the status quo. Instead, the increase in the audit rate ranges between 0 and 2 additional percentage points in intervals of 0.5 percentage points (i.e., a total of five values). Hence, we have a total of twenty additional cases to evaluate per successful outcome from part 1. The trajectories that are shown in Figure 22.7 reveal that although the compliance level is not very sensitive to the new policy, the dynamics of the marginal debt (including if and when it is decreased back to zero) is very sensitive to the chosen policy.

In our analysis of part 1, we considered a total of 50 cases for each of our 48 policy experiments, totaling 2400 cases. Out of these 2400 cases, 386 cases shown reached our target compliance while at the same time producing a marginal debt. In part 2 we consider 20 policy experiments for each of these 386 cases, totaling 7720 cases. Out of these cases, 2255 (i.e., 29%) had a successful outcome and belong to the set of desirable policies.

Figure 22.8 provides a heat map showing a summary of the analysis of our 2255 desirable cases. The figure mirrors the Figure 22.4 shown previously which analyzed the starting 386 cases from part 1. As before, each cell corresponds to a given policy setting from our part 1 exploration. Cells shown in red indicate that it contained none of the desirable policies found in part 2. For example, as can be seen in the cell at the lower right corner of Figure 22.4, many policies considered described by this cell during part 1 reached our compliance target. However, on average it required 24 years to reach the compliance target and due to the very aggressive reduction in the tax rate, it produced an average marginal debt of $7.2 trillion over this period of time. When we
then compare this to the results in Figure 22.8 we see that there are no part 2 desirable policies bounded by our exploration range which can reduce this large marginal debt to zero.

Cells shown in lighter colors ranging from orange to green indicate that they contain an increasing number of part 2 desirable policies. For these cells we report three average quantities:

1. The number of years needed to decrease the marginal debt back to $0.
2. The tax rate multiplicative factor applied during part 2.
3. The additional audit rate applied during part 2.

Figure 22.8 suggests if the tax rate is increased, the audit rate needs to remain high and increased by an additional 1 pp in order to maintain the compliance level at 95%. To reduce the accrued marginal debt from the dynamics in part 1, the taxes need to be essentially brought back to rates that were applied at the status-quo.

Comparing these findings to those shown in the first row of Figure 22.3, it becomes visually evident that these results suggest that policies that combine tax rate reductions with an increase in the audit rate to achieve higher compliance do not offer strong advantages compared to a direct increase in the audit rate without a tax rate reduction. There is little indication that higher compliance can be reached sooner by such combined policies involving a part 1 and part 2 component. As opposed to analyzing averages, case-specific analyses reveal some advantages in using such combined policies but these are small and are not worthwhile.

---

<table>
<thead>
<tr>
<th>Audit Rate Additive Factor (pp)</th>
<th>4yr</th>
<th>19yr</th>
<th>7yr</th>
<th>8yr</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.5</td>
<td>108%</td>
<td>108%</td>
<td>108%</td>
<td>108%</td>
</tr>
<tr>
<td>2.75</td>
<td>1%</td>
<td>1%</td>
<td>1%</td>
<td>1%</td>
</tr>
<tr>
<td>3</td>
<td>110%</td>
<td>110%</td>
<td>110%</td>
<td>110%</td>
</tr>
<tr>
<td>3.3</td>
<td>1%</td>
<td>1%</td>
<td>1%</td>
<td>1%</td>
</tr>
<tr>
<td>4</td>
<td>31yr</td>
<td>53yr</td>
<td>39yr</td>
<td>35yr</td>
</tr>
<tr>
<td>5</td>
<td>99yr</td>
<td>114%</td>
<td>114%</td>
<td>113%</td>
</tr>
<tr>
<td>6</td>
<td>1.1%</td>
<td>11%</td>
<td>1%</td>
<td>1%</td>
</tr>
<tr>
<td>7</td>
<td>NaN</td>
<td>NaN</td>
<td>NaN</td>
<td>NaN</td>
</tr>
<tr>
<td>8</td>
<td>NaN</td>
<td>NaN</td>
<td>NaN</td>
<td>NaN</td>
</tr>
</tbody>
</table>

Figure 22.8: Heat map summarizing the findings of our part 2 exploration. Each cell provides details of the policy experiment.
Summary, Discussion, and Conclusion

23 Summary and Discussion

We built an agent-based computational simulation model of income tax evasion to explore a range of different policies that can increase taxpayers’ compliance. Within the simulation, taxpayers’ compliance behavior changes through an adaptation process based on their past experiences with audits and tax evasion penalties, their perception of the fairness in taxation rates and social interactions with people in their social networks. The model was designed to explore policies that could lead to cascading amplification effects and potentially to societal transformations regarding tax compliance that can persist over time. This is an ambitious goal since compliance behavior involves complex nonlinear responses and interactions whereby taxpayers self-organize and adapt to new environments that they, in part, helped to form. Building on our past work, we linked our simulation model with a tailored behavioral and social network survey and develop a survey-and-simulation modeling framework to describe tax compliance behavior. The survey was used to inform the model about taxpayers’ perceptions of tax fairness and the risk of being audited by the IRS. This approach is novel: We have previously developed a similar survey-and-simulation modeling framework to model various complex and adaptive systems relating to health policy. However, for the first time, we have applied this approach to describe a non-health complex and adaptive socio-economic system affected by government policies. Moreover, our model was informed from multiple secondary data sources including census data, IRS data and network data from the Network Dynamics and Simulation Science Laboratory (NDSSL).

23.1 Summary of our survey design and analyses and secondary data used

Soon after the model was conceptualized, first versions of the model were developed and tested. This allowed us to get an early sense of how the model performed and provided first indications of model sensitivity, and provided the opportunity to carefully consider data needs and design our survey questions to best inform the model. Surveys on a sensitive topic like tax evasion behavior are hard to design and analyze and this took a prolonged part of the project’s time. Analyses of our survey data were used to inform the three general components of our ABM. Namely, (i) the taxpayers’ attributes and distributions that affect tax compliance behavior; (ii) the way taxpayers’ mix socially and discuss tax-related topics; and (iii) the ranges for model parameter values that affect the taxpayers’ compliance behaviors in the model. Our aim was to provide estimates with uncertainty bounds of parameter values and attribute distributions describing the taxpayers’ behaviors. The law of diminishing returns sets in very quickly when estimating parameter values for the purpose of informing the model. Thus, our priority was to provide good-enough estimates and allowed the model calibration to help narrow parameter estimates. Sensitivity analysis of our model results helped to guide in determining which parameter estimates and survey data analyses required more attention. The methods we used to inform the taxpayers’ attributes included a combination of analyzing and fitting smoothed distributions and running a set of regression models describing the perceptions of the respondents. Our ABM considers how taxpayers mix socially. Therefore, we had to assign the taxpayers’ attributes in a way that preserved important social mixing patterns such as employment status, income, and tax-related perceptions. We used a scaled-down version of the network for Portland OR provided by the Network Dynamics and Simulation Science Laboratory (NDSSL). This network data contains important attributes such as family income and the number of family members. This data together with regression mod-
els allowed us to assign the additional taxpayers’ attributes on the network including whether a taxpayer is self-employed or not and the various baseline fairness and risk perceptions relating to taxes and IRS audits. The way these attributes were assigned preserved the fitted smoothed distributions and the level of assortative mixing found in our analyses of the survey data. We computed elasticities and ran panel-level regression models of the responses to the hypothetical scenarios to understand how perceived compliance changes with perceived effective tax rates, audit rates, and penalty rates. In particular, we used the panel-level regression models to construct a multi-variate S-curve response function that directly informed our ABM. This S-curve describes how taxpayers propensity to initiate in tax evasion changed with respect to the perceptions of their effective tax rate and on the perceptions of the risks of being audited and penalized by the IRS for tax evasion. This analysis represents a novel approach to informing the behavioral attributes of agents in an ABM.

According to the results our survey analyses, taxpayers have very high perceived audit rates. The average perceived audit rate was over 20%. We think that most people perceived that simple checks or letters from the IRS are a tax audit. The perceived audit rate does not decrease based on experiencing a recent audit. Instead, respondents in our survey believe that the risk of being audited strongly depends on the proportion of income reported to the IRS. These findings were used to help inform behavioral mechanisms describing how taxpayers change their perceived risk of being audited via adaptation. However, we note that this is based on respondents self-report and does not necessarily describe what a taxpayer would do after experiencing an IRS audit, or once they initiate in underreporting on their income. To account for this uncertainty, we also relied on other estimates from the literature to estimate the ranges of values describing these mechanisms. Respondents were also asked to estimate their perceived effective tax rate and the perceived penalty rates for being caught by the IRS for underreporting. Respondents perceived their effective tax rates to be rather high and closer to what their marginal tax rate would be. The perceive penalty rates were also high. We asked respondents to estimate perceived underreporting rates in taxpayers similar to themselves under the baseline scenario and a set of hypothetical scenarios that included changes in the tax rate and risks and penalties of being audited and caught for tax evasion. Using a fixed-effects panel regression models we estimated the elasticities of the perceived underreporting rates on the effective perceived tax rate, perceived audit rate and perceived penalty rates. We found statistically significant a positive elasticity on the effective tax rate of near one and negative elasticities with perceived audit rate and penalty rates with a magnitude lower than one. Using a similar approach, we estimated our S-curve describing the dependence of the perceived evasion rates on these three perceptions. This was used to inform the model.

Our survey was also used to estimate network and media effects on tax-related perceptions. Our findings showed that respondents believe that people like them are more compliant than the typical taxpayers in the US. Respondents had significantly lower perceived underreporting rates among people like them. However, if exposed to news that reported a significantly large average rate of tax evasion in the population, respondents believe that taxpayers like themselves would respond by underreporting at the same or at an increased level as the taxpayers in the population. This analysis helped inform how taxpayers in our model may respond to media announcements regarding population-level average compliance rates. Respondents also reported interacting socially and discuss tax-related topics with their social contacts (i.e., alters). On average our respondents talk to little over a quarter of their alters about taxes on a yearly basis. This suggested a strong component network effects. Interestingly, we did not, however, find evidence that the percent of alters who have been audited is associated with perceived underreporting rates. The same was true for the percent of alters with whom the respondents talk about taxes. However,
the percent of alters who are self-employed is positively associated with perceived underreporting rates. This is especially true for perceived underreporting among people like themselves. We used these findings, regression analyses and iterative methods to inform the way taxpayers’ mix socially and discuss tax-related topics for our model. We found that when making evaluations regarding fairness and risks of being audited by the IRS, respondents primarily rely on their own personal perceptions and experiences regarding taxes and less so on social interactions and exposure to media. The importance of social interactions and exposure to media on these evaluations carried approximately equal weight and taken together about 40% of the weight.

23.2 Summary of model verification, validation, and calibration

After informing the model using the analyses of our survey data and of secondary data sets, we proceeded with additional model verifications, external validation, uncertainty and sensitivity analyses, and model calibration. Model validation was based on using expert opinion of model design and outputs, and test whether it could reproduce known stylized facts and summary statistics describing tax compliance behavior and its aggregated outcomes that were not used to inform the model. For example, the model reproduced the U-shaped distribution describing individual-level compliance behavior, generally high levels of persistence taxpayers’ compliance behavior, tax gaps that are not wildly different than those that naturally occur and overall average audit rate costs. Since the IRS does not fully disclose their rules and processes for selecting taxpayers for an audit, the latter validation indicated that our description of the plausible rules implemented in our model and describing the IRS selects criteria are not unreasonable.

Through our uncertainty and sensitivity analysis, we learned how the model responds to the different inputs. Unsurprisingly, we found that overall tax compliance as measured by the tax gap is sensitive to the audit rate, the tax rate and how taxpayers change their compliance behavior after being penalized. The latter depends on the amount of unpaid taxes and penalties an evading taxpayer is required to pay, which depends on the penalty rate and the number of years s/he was caught evading. In our sensitivity analysis, we varied the additional audit rate in a range of \( \pm 0.5 \) percentage point and varied the effective tax rates by a multiplicative factor in a range of \( 100 \pm 8\% \). Our analysis showed a marginally higher sensitivity to the tax rate. However, the range assumed for the audit rate here is much smaller than the range used for the tax rate which suggests a higher sensitivity to the audit rate than to the tax rate when we consider narrower variabilities in the tax rate. This observation helps understand findings from our policy experiments. Other key drivers include (i) the strength (i.e., weight) that taxpayers are influenced by their tax-related social network contacts compared to their own experiences and changing perceptions and (ii) how strongly taxpayers discount past experiences, outcomes, and perceptions compared to more recent experiences. Interestingly, although the strength of social influences is a key driver of taxpayers compliance behavior, our model suggests that the average number of tax-related social contacts is not a strong determinant for tax compliance at the aggregated level. Another interesting finding is that although the audit rate and the way taxpayers change their compliance behavior after being penalized affect the overall tax compliance behavior, it does not strongly affect the U-shaped distribution describing individual-level compliance behavior and the proportion of the hideable income they report. The reason for this counterintuitive outcome is that an increase in the audit rate does not mean a change in the audit strategy. Therefore, increasing the audit rate in our model primarily results in selecting the same type of taxpayers more frequently rather than diversifying the taxpayers being audited. Consequently, the U-shaped distribution is only marginally affected.

A particularly laborious step was model calibration. During calibration, we found that our model produces many case runs with tax gaps close to our calibration target (i.e., within 0.5%) and
can also produce many case runs which reproduce our target U-shaped distribution of tax compliance behavior. However, the number of cases that can accurately reproduce both these targets (i.e., their overlap or intersection) was relatively small. A principal component analysis revealed a tension between satisfying both these two main calibration targets and revealed tensions between other calibration targets which aligned more with either the tax gap or the U-shaped distribution target. Based on a prolonged iterative trial-and-error approach, we chose calibration weights that revealed to be consistent with our principal component analysis and provided what we think is a good compromise between the two key targets. However, a larger latin hypercube experimental design for our calibration could have helped identify a larger set of cases runs that better satisfy both these two calibration targets. There are iterative calibration methods that use machine learning or inputs by expert opinion (e.g., Bayesian melding approaches) to guide and improve the calibration. These will be explored in further studies. Moreover, although we calibrated the model by exploring parameter space we did not explore alternative ways of assigning taxpayers’ attributes. As can be inferred by inspecting Figure 17.13, swapping attributes values between taxpayers such as their income, their proportion of hideable income and their self-employment attribute in a way that preserves consistency and overall statistical distributions may be an effective way of easing the tension between our calibration targets.

23.3 Summary of hypothetical "what-if" scenario experiments

We carried out a set of experiments that considered hypothetical "what-if" scenarios to explore what our model predicts for these cases. In these experiments, we started the system using a set of initial calibrated states that describe the baseline (i.e., status-quo) settings and allowed the model to reach a new stationary state dictated by the policy specifications. We then compared compliance levels and distributions in these new stationary states with respect to those describing the status-quo. These experiments considered: (i) a decrease in the Gini coefficient describing a more equitable income distribution that is similar to the one for Sweden; (ii) a decrease in the self-employment rate; and (iii) a simple change in the audit strategy making IRS audits irregular and intermittent over the years.

Our model suggests that making the incomes distribution more equitable without changing the audit strategy used by the IRS could lead to an increase in the tax gap. This counter-intuitive population-level finding was confirmed on detailed inspection of compliance behavior at the individual-level. The reason for this effect is that numerous non-compliant lower-income taxpayers that hardly ever get audited have larger incomes in the new scenario, but continue to be a low priority for IRS audits. Concurrently, less numerous high-income taxpayers have less income to report to the IRS but remain a high priority on the IRS list of taxpayers to audit. The net effect is that the tax-gap marginally increases.

Our model also suggests that decreasing the proportion of self-employment taxpayers, who are characterized by having all of their income being potentially hideable to the IRS will at first increase the tax-gap before decreasing it as we approach 0% self-employed. We found that reducing the proportion of self-employed taxpayers from 16% to 10% would increase the tax gap. When we further decrease the proportion of self-employed to 5% the tax gap remains larger than that for 16% but decreases compared to the case with 10% proportion of self-employed. We verified that at 0% proportion of self-employed the tax gap is significantly decreased compared to the status-quo scenario 16% proportion of self-employed. This result is also a counter-intuitive population-level finding which was understood by inspecting individual-level behaviors. It is explained by a marginal decrease in compliance behavior of the non-self-employed taxpayers. These taxpayers are more numerous under the new scenario and at the individual-level experience less
frequent audits. This leads then to perceived a marginally smaller risk of being audited which makes them less compliant. Under the new scenario, the self-employed taxpayers are less numerous and continue to be a high priority on the IRS list of taxpayers to audit. Consequently, they experience more frequent audits and their compliance behavior increases. However, at the aggregated or population-level the reduction in compliance of the non-self-employed is stronger and is not balanced by the increased compliance of the self-employed. The net effect is a reduction in compliance.

We also explored a hypothetical scenario whereby the IRS audit the taxpayers regularly and in intermittent years. The idea here was to test whether the increased level of uncertainty of IRS audits induces an increased level of fear and higher compliance rates in taxpayers in our model. The intermittency of IRS audits would also lead to more frequent media feedback effects in our model, where the media reports about unusually high and different IRS auditing activities back to the taxpayers. However, results do not suggest increased compliance. Compliance levels are largely unaffected and if anything we see a decrease in compliance. To understand this result we need to compare the timescales between the periods of audit years the time scales used by the taxpayers when they evaluate their tax-related experiences and adapt. Since in our model the latter are relatively short time scales, large periods of IRS inactivity would lead taxpayers to quickly discount the risk of being audited.

23.4 Summary of tax reform policies and changes in the audit rate

We tested the effects of the recent tax reform that introduced a new tax schedule with reduced marginal tax rates. As before, to test this policy we started the system using our status-quo settings and allowed the model to reach a new stationary state before making comparisons. Our model suggests that on average the tax reform will reduce the tax gap but only by a small amount of about half a percentage point. Unless taxpayers generate more income due to this tax cut, our model suggests that there will be a reduction in tax revenues of over $200 billion. This estimated reduction is in agreement with other recent estimates. This finding suggests that reducing the tax rate without increasing the audit rate and assuming that it doesn’t increase GDP and taxpayer’s income will only marginally decrease the tax gap but will strongly increase the government’s yearly budget deficit.

We found that there are strong direct and indirect benefits of increasing the audit rate. If the overall audit rate is increased by over 2 percentage points from its current value of just under 1% to 3%, the increase in IRS costs for performing these additional audits would be covered by the direct increases in revenues from catching the non-compliant taxpayers and recovering unpaid taxes. However, this finding assumes that once taxpayers are caught for unpaid taxes they pay this what is due to the IRS without problems or delays. There are also strong indirect benefits of increasing the audit rate to 3%. Our model suggests that the tax gap could be reduced by 6 percentage points on average and this would lead to recovering more than $100 billion in tax revenues due to the increased level of compliance. These indirect benefits increase as we further increase the audit rate. Our experiments suggest that the marginal increase in tax revenues is larger than the marginal increase in the audit cost when the audit rate is below 13%. Such a large increase in the audit rate would essentially reduce the tax gap to a little less than 1%. However, our model does not consider the effects of the increased burden and frustration on taxpayers that an audit rate of 13% would likely produce. Such a large and steep change in the audit rate without changes in operations and bureaucracy could be socially unacceptable and lead to likely lead to widespread malcontent. Therefore, although these are interesting finding we do not suggest nor recommend such a large increase in the audit rate. However, based on our findings we think
increasing the audit rate to 3% should be considered.

23.5 Summary of policies that combine changes in tax and audit rates

The last set of policy experiments explored in this report combined changes in audit rates and tax rates to explore the dynamics of both the compliance and of the marginal debt. The aim of this exploration was to find socially acceptable policies that can achieve an increased compliance reaching a target level of 95% while leading to economically manageable marginal debts over the years of the policy. We then explored what policies would be able to pay back the accrued marginal debt while sustaining the achieved higher compliance level. Our exploration used a large range tax rate changes and audit rate changes. However, the exploration used discrete and coarse intervals within these ranges.

Our findings showed that in order to reach a voluntary compliance rate of 95%, the audit rate will need to increase by at least 2.75 percentage points to just under 4%. The encouraging finding is that such a policy would produce a marginal surplus over the years rather than a marginal debt, even if the overall tax rate were reduced by 5% compared to the rate assumed in the status-quo (i.e., those for the year 2016). However, further reductions in the tax rate do not necessarily help in achieved the higher compliance rate sooner and produce large marginal debts which would need to be repaid later. Tax compliance cannot be increased rapidly with a tax reduction unless it is accompanied by an increase in the audit rate. The reason this occurs in our model is as follows. Decreases in the tax rate lead compliant taxpayers to be less likely to initiate in tax evasion behavior. However, taxpayers that have been underreporting most or all their hideable income for years do not become more compliant because of the reduction in the tax rate. Instead, they can only become compliant if caught by the IRS. Consequently, reaching the target compliance mainly depends on the audit rate increase. The reduction in the tax rate only slows down the rate at which taxpayers initiate or relapse back into non-compliance behavior.

A policy that concurrently introduces an increase in the audit rate to 4% and a reduction in the tax rate of over 5% would reach the target compliance level of 95% but would lead to additional deficits and an accrued debt which would then need to be repaid with painful increases in the tax rate later on. Moreover, due to this latter increase in the tax rate, in order to sustain the compliance level at 95%, the audit rate would need to be increased further by 1 percentage point. Therefore, policies that decrease the tax rate can achieve higher compliance but rapidly become prohibitively expensive as they accrue additional debts that will not be paid off for many decades, even by the increased revenues generated by the higher compliance rate. Our findings are partially grim but not surprising and suggest that in order to increase compliance, tax rate reductions should be introduced slowly over time depending on the increases in the audit rate.

24 Limitations

"All models are wrong but some are useful" is a quote by statistician George Box. Indeed, the idea that a model can make accurate predictions does not necessarily mean that the assumptions and the fundamental principles governing the model are correct. A classic example of this is

83A caricature of our finds is that the stick (i.e., increase audit rates) is more effective than the carrot (i.e., reduced tax rates), and the carrot can only be provided at a rate that depends on how effective the stick is over time. Having carrots too much and too soon causes a stomach ache to develop and whose remedy is an additional increase in the stick.
As first confirmed by Galileo we now know that the model is wrong. The simpler heliocentric model first proposed by Copernicus provided the same predictions as the Ptolemaic system with fewer ad hoc assumptions. It can be said that the Copernican heliocentric model follows a problem-solving principle known as Occam’s razor which states that when multiple competing hypotheses are equal in all other respects, we should select the hypothesis that introduces the fewest assumptions and postulates and that contains the fewest entities. However, this is not to say that models that are more complex and complicated are necessarily wrong. Rather, we should carefully consider the purpose and the level of detail that a model is asked to explain and predict before using the principle of Occam’s razor. For example, Newton’s mathematical description of motion and of gravity has proved to be sufficiently accurate for most practical level physics. However, although it describes motion and gravity it does not explain it. Einstein’s theory of General Relativity, a more complicated mathematical theory is needed to explain motion under gravity. Einstein’s opinion is that more complicated models "should be considered only if there exist physical-empirical reasons to do so."

The purpose of our ABMs was to simulate both the micro-behavior and macro-behavior of the complex socio-economic system describing tax compliance. This is an ambitious task. In our model conceptualization we considered many of the key mechanisms that affect tax compliance behavior and relied on literature review, expert opinion and data from our surveys to inform the model. However, we postulated micro-level behavioral mechanics and made assumptions and simplifications. Here we list some of these and describe their limitations. We note however this list is not exhaustive.

1. It is known and confirmed by our survey that many taxpayers rely on Certified Public Accountants (CPAs) and tax software to compile their tax returns. Yet this was not included in our model. Taxpayers that share the same CPA may have similar reporting behaviors even if they do not know each other. This is an additional network effect which is not included in our model and is left for future work. In addition the IRS may have audit strategies that select taxpayers based on a common CPA.

2. Whether a taxpayer receives a tax refund or is asked to pay a balance due in their tax returns influences his/her compliance behavior. This has been included in the model. However, as a simplification, the model does not track the amounts of refunds, balance due and deductions. Therefore, just the sign and not the magnitudes of these amounts affect the taxpayer’s compliance behavior in our model.

3. As long as the marginal tax rate and tax brackets do not change, our model assumes that the effective tax rate applied to a taxpayer does not change as well. However, in reality the effective tax rate depends on the amount of income that a taxpayer reports to the IRS, and this changes from year to year based on his/her level of compliance. Hence the effective tax rate applied to a taxpayer should dynamically change even when tax rates are unaltered. We tested the effects of including this dynamics and found that it qualitatively produced similar distributions describing taxpayers’ compliance behaviors.

4. Our model focuses on tax evasion (which is illegal) and not on tax avoidance (with is legal).

---

84The Greco-Roman mathematician Ptolemy created an extremely complicated mathematical model for the motion of planets in the solar system which assumed that the Earth is at the center of the system and the planets moved in loop-type paths called epicycle. Ptolemy’s geocentric model reliably and accurately predicted the future trajectories of the planets.

85For the purpose of space missions Newtonian mechanics has proven to be sufficient.
However, there is a fine line between the two behaviors. Moreover, taxpayers may not be aware that they are evading rather than avoiding paying taxes on various sources of income.

5. Our model assumes that all taxpayers file tax returns and some intentionally underreport all or significant proportions of their hideable income. The model does not consider taxpayers that do not file tax returns at all but may have taxable sources of income.

6. Corporate tax evasion is not considered. Our model only focuses on individual income taxes. However, individual income taxes can be affected by their perceptions of what corporations do and how compliant they are in paying corporate taxes.

7. As a simplification, the model does not assume vital dynamics. Essentially, taxpayers in our model do not die and new ones are not born. The network describing their interacting does not change over time. This simplification, has been tested in other previous ABMs we developed. As long as the timescales describing the discount rates of new experiences, perceptions, and behaviors is much smaller than the time scale describing the taxpaying lifetime of an individual, this simplification does not have a strong effect on model results.

8. Tax morale is modeled entirely on a taxpayers effective tax rate. However, tax morale depends also on additional things such as perceptions and trust on how well the government is spending taxes collected from the taxpayers, political alignment and services provided. These effects are implicitly modeled by assuming that taxpayers have different tax rate thresholds that describe the tax rate which they would consider before initiating in tax evasion. The heterogeneity and distribution of this threshold was informed by our survey.

9. For each case run the weights that taxpayers use for personal evaluations, the effect of social interactions and media effects is assumed to be the same across all taxpayers. Although these weights vary from case to case, within any one case run all taxpayers use the same set of weights. The reason we chose to do this was to make sure that we could vary these weights from case to case and see how they affect the dynamics. If we assumed heterogeneity of these weights within each case run, we would have required more complicated ways of varying these from case to case making it hard to evaluate their effects via a standard sensitivity analysis.

10. Many of the taxpayers’ attributes are assumed to be constant throughout the dynamics. For example, the thresholds that describe the tax rate which taxpayers consider before initiating in tax evasion and the evaluation weights. However, as tax compliance changes at the aggregated level, these attributes describing taxpayer behaviors could also change.
Appendicies

A  Experience Discounting and the EWMA process

Our experience discounting mechanism follows a similar approach that we have used in other past ABMs. We start with the expression

\[ V_{t+1} = sV_t + \Delta_t \]  \hspace{1cm} (A.1)

where \( \Delta_t \) is our experience in year \( t \) that is bounded between 0 and 1 and \( s \) is the discounting parameter which is also bounded between 0 and 1. When the outcome of a given decision in year \( t \) was positive agents re-enforce making the same decision in the future and thus \( \Delta_t = 1 \). If \( s = 1 \), then \( V_t \) represents the tally of the number of times the decision lead to a positive outcome. We then normalize \( V_{t+1} \) by \( N_{t+1} = sN_t + 1 = (1 - s^{t+1})/(1 - s) \) to produce a probability \( w_{t+1} \). Therefore, \( w_{t+1} \) can be expressed as

\[ w_{t+1} = V_{t+1}/N_{t+1} = (N_t/N_{t+1}) \cdot [sN_t + \Delta_t/N_t]. \]  \hspace{1cm} (A.2)

The normalization factor \( N_t \) represents the maximum value \( V_t \) could be if \( \Delta_y = 1 \) for all years \( y \leq t \) and is simply the sum of a geometric progression. However, notice that \( \Delta_t \) doesn’t necessarily need to be bounded in \([0, 1]\) as long as we properly normalize \( V_{t+1} \). For example, if taxpayer \( i \) experiences a particularly negative outcome such as being audited and penalized for tax evasion, it could be a very salient experience that it essentially wipes out the importance of any prior experiences. In this case the normalization \( N_{t+1} \) is no longer simply equal to \( (1 - s^{t+1})/(1 - s) \)

For the cases when \( \Delta_t \) is bounded in \([0, 1]\) we can show that this procedure is essentially an exponentially weighted moving average (EWMA) \([147]\), which can be expressed as

\[ w_{t+1} = rw_t + (1 - r)y_t. \]  \hspace{1cm} (A.3)

By equating equations \[ A.2 \] and \[ A.3 \] we find that \( r = sN_t/N_{t+1} = s(1 - s^t)/(1 - s^{t+1}) \) and that \( (1 - r)y_t = \Delta_t/N_{t+1} = (1 - s)\Delta_t/(1 - s^{t+1}) \). Thus in the limit of large \( t \) the ratio \( N_t/N_{t+1} \rightarrow 1 \) and we have that (i) \( r \rightarrow s \) and (ii) \( y_t = \Delta_t \). Thus, the two processes are equivalent.

A.1  Example EWMA process that are relevant to our model

Let’s consider a few examples

1. If we assume \( \Delta = 0 \) we have that \((w_{t+1} - w_t) = (s - 1)w_t \) and thus in the continuous limit this can be expressed as

\[ \dot{w}_t = -(1 - s)w_t \]  \hspace{1cm} (A.4)

leading to an expression \( w_t = w_0e^{-(1-s)t} \). Thus, according to the formal definition of the discounting rate of the exponential process this is equal to \( 1 - s \). Moreover, for the continuous limit the half-life is equal to \( \log 2/(1 - s) \). Thus if \( \Delta = 0 \), over time \( w(t) \rightarrow 0 \). Translating this solution to the discrete case we obtain what we expected and \( w_t = w_0e^{-(1-s)t} \Rightarrow w_{t+1} = w_0e^{-1} \sim w_1[1 + (s - 1) + (s - 1)^2/2! + (s - 1)^3/3! + \ldots] \). Since, \( s \) is positive and less than 1, we can approximate the limit of this sum with \( sw_1 \) where the half-life is \( -(\log 2)/\log(s) \).
2. If instead we assume that $\Delta$ is the same constant every year and using equation A.3 we see that

$$w_{t+1} - w_t = (s - 1)\Delta,$$

which leads to

$$w_t + (1 - s)\Delta = (1 - s)\Delta.$$  \hspace{1cm} (A.5)

Using the integrating factor $e^{(1-s)t}$ we find that $w_t = \Delta \{1 - e^{(s-1)t}\}$. Thus, since $s < 1$ we have $w_t \rightarrow \Delta$.

3. If we assumed that when $\Delta < 1$, the value of $\Delta$ is stochastically assigned to be either of two values, $\chi$ (bounded in $[0, 1]$) or 0 with probabilities $w_t$ and $1 - w_t$ respectively, then the expected value for $\Delta$ is $w_t \chi$. Thus, the expected value for $w_{t+1}$ (denoted by $\bar{w}_{t+1}$) is given by

$$\bar{w}_{t+1} = s\bar{w}_t + (1 - s)\bar{w}_t \chi$$

which leads to $\bar{w}_{t+1} - \bar{w}_t = -w_t(1 - s)(1 - \chi)$. Replacing $w_t$ with $\bar{w}_t$ and solving the differential equation obtained in the continuous limit we get

$$\bar{w}_t = \bar{w}_0 e^{-[(1-s)(1-\chi)t]}.$$

Hence, although $\bar{w}_t \rightarrow 0$ the decay is slower and characterized by a half-life time scale of $\log 2/[(1 - s)(1 - \chi)]$. Translating the solution from the continuous to the discrete time case we have

$$\bar{w}_t = \bar{w}_0 e^{-[(1-s)(1-\chi)t]} \Rightarrow \bar{w}_{t+1} = \bar{w}_1 e^{(1-s)(1-\chi)} \sim w_t[1 + (s - 1)(1 - \chi) + (s - 1)^2(1 - \chi)^2 + \ldots] \sim [1 + (s - 1)(1 - \chi)]w_t = [s + \chi(1 - s)]w_t.$$

Thus the new modified discount factor is $[s + \chi(1 - s)]$ and the discrete time half-life is equal to $-\log 2/\log[s + \chi(1 - s)]$. This example, illustrates the change in the decay time scale due to the chosen stochastic update for the personal evaluations described in Section 7.1 and in equation 7.5 for the simple and particular case where $\nu = 1$ and for $F(T_i[1], \delta_i[1])$ equal to a constant $\chi$.

### A.2 Modeling the effects of a salient outcome on the EWMA process

The experience $\Delta_t$ normally ranges between 0 and 1. However, certain outcomes like being audited and penalized by the IRS will be a particularly salient event and a penalized taxpayer will recall and factor this event in his/her decision-making for a long time. Therefore, experiences of being penalized may strongly diminish the importance of previous experiences in terms of determining the future compliance behavior of the taxpayer. We label the experience of being penalized in year $t$ as $\hat{\Delta}_t$ and its value can range from 0 to well above 1. In our model description, $\hat{\Delta}_t$ represents $\Delta_{penalty}$ and is given by the proportion between the amount of past unpaid taxes the taxpayer owes the IRS plus penalties, and the amount of annual taxes s/he normally owes the IRS.

For example, consider the case whereby $\hat{\Delta}_t = 1$. This occurs when the taxpayer penalties plus past unpaid taxes is equal to amount of annual taxes which s/he normally owes the IRS. Let’s further consider the case whereby the taxpayer propensity to report their full income is zero. Hence, in our model this would occur because $\bar{V}_t$, and thus $w_t$ equal to 0. We assume that under this scenario the taxpayer’s $w_t$ value increases from 0 to a given constant value $\bar{V}$ in the following year (i.e., $w_{t+1} = 0$). Although in our model taxpayers who are caught and penalized by the IRS are assumed to become compliant in the following year, their propensity to remain compliant does not generally increase to 100%. In this example case, the propensity to comply increases from 0 to the value $\bar{V}$. Thus,

$$w_{t+1} = \frac{\bar{V}_{t+1}}{\bar{N}_{t+1}} = \frac{s\bar{V}_t + \phi \hat{\Delta}_t}{s\bar{N}_t + \phi \hat{\Delta}_t} = \frac{sw_t\bar{N}_t + \phi \hat{\Delta}_t}{s\bar{N}_t + \phi \hat{\Delta}_t},$$  \hspace{1cm} (A.6)

and since in this case $\bar{V}_t = 0$ and $\hat{\Delta}_t = 1$ we have

$$w_{t+1} = \frac{\phi}{s\bar{N}_t + \phi} = \theta,$$  \hspace{1cm} (A.7)
where the constant $\phi$ is determined by the particular choice of the value $\theta$ and is given by $\phi = sN_t/\theta(1-\theta)$. We use this example case to define the way the value of $\phi$ is calculated. So for the general case whereby $w_t > 0$ and the taxpayer is audited and penalized in year $t$, his/her new value for $w_{t+1}$ is evaluated as follows

$$w_{t+1} = \frac{sw_tN_t + \phi\hat{\Delta}_t}{sN_t + \phi\Delta_t} = \hat{s}_t w_t + (1 - \hat{s}_t) \cdot 1,$$

(A.8)

where $\hat{s}_t = s/[s + (1-s)\phi\Delta_t]$ for the case when $N_t$ has reached its large time limit of $(1-s)^{-1}$. Thus, we can see that when a penalty occurs we retain the EWMA update but the effective discount parameter $\hat{s}_t$ is much lower than $s$ and this has an effect of strongly diminish (i.e., "washing away") the importance of previous experiences with respect to the effect of the penalty. The strength of the diminishing effect (i.e., $\hat{s}_t$) is determined by both the penalty value $\hat{\Delta}_t$ and our choice for $v$ and hence the value of $\phi$.

Returning to the case that $\hat{\Delta}_t = 1$ and $w_t = 0$, we now reset the time counter to zero and use $t$ to denote the number of years since being penalized. Equation A.7 states that $w_1 = \theta$. In our model $w_{t+1}$ is normally used to provide the probability that a compliant taxpayer initiates in under-reporting. However, we enforce that a taxpayer who is caught underreporting and is penalized will become compliant in the subsequent year. Thus, the probability to initiate in under-reporting a year after being penalized is zero. Initiation or relapse back to tax evasion behavior is only possible in our model in years $t > 1$. The expected time for the taxpayer to relapse and being to under-report again can be expressed as

$$\langle t_{\text{init}} \rangle = \sum_{t=2}^{\infty} tw_t.$$

(A.9)

In our model, $w_t$ is influenced by the new value for $\Delta_t$ which factors in the changes in perceived risks of being audits and penalized. After being audited and penalized, the value of $\Delta_t$ will have increased. However, assuming that in the subsequent years the values for $w_{t+2}$, $w_{t+3}$, $w_{t+4}$ etc. do not change and remain equal to $\theta$ we can estimate the expected time that the taxpayer will relapse and being to under-report again. This is simply equal to

$$\langle t_{\text{init}} \rangle = (1-\theta)^{-1}.$$

(A.10)

A.3 Time scales in the EWMA process

In the previous sections we have (i) obtained an expression for the mean time to initiate in tax evasion used and (ii) found an expression for the half-life for the stochastic update used for the personal evaluations. If we assume that once a taxpayer reports 10% or less of his/her income s/he is effectively close to full evasion, we can estimate a generalized time scale $\bar{\tau}$ for the case whereby a taxpayer relapses back to full evasion given by

$$\bar{\tau}_{1/10} \sim \{ \langle t_{\text{init}} \rangle + \langle t_{\text{decay}} \rangle \} = \{ \langle t_{\text{init}} \rangle - \log 10 / \log[s + \chi(1-s)] \}$$

(A.11)

86In a previous version of our model we set $\phi = N_t$ and thus the value of $\theta$ was not a free parameter but rather was fixed by the value of $s$ and equal to $\theta = (1+s)^{-1}$. Allowing for $\theta$, and hence $\phi$ to vary allows us to scale and choose the strength of the influence of being penalized on the propensity to comply with respect to past evaluations, which generally do not involve being penalized.
The expression for $\bar{\tau}$ considers all the time-scale parameters that enter our model that affect compliance behavior (except for the generation lifetime). This time-scale is useful and can be compared to the audit rate used by the IRS assumed in our model. This is a crude estimate of the time scale where we assumed that $\chi$ is a constant and the tax evasion initiation probability is also a constant and doesn’t change base the changing risk perceptions. The real-time scales highly depend on the taxpayer’s attributes and compliance and risk perception trajectories. Nevertheless, the expression for $\bar{\tau}$ still allows us to combine parameters and get an idea of the generalized time-scale that affects our model. Table A.1 provides some calculated examples of the $\bar{\tau}_{1/10}$ time scale.

<table>
<thead>
<tr>
<th>$\tau$</th>
<th>$\theta$</th>
<th>$\chi$</th>
<th>$s$</th>
<th>$\langle t_{\text{init}} \rangle$</th>
<th>$\langle t_{\text{decay}} \rangle$</th>
<th>$\bar{\tau}_{1/10}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.71</td>
<td>1.00</td>
<td>6.60</td>
<td>7.60</td>
</tr>
<tr>
<td>2.00</td>
<td>0.58</td>
<td>0.00</td>
<td>0.71</td>
<td>2.40</td>
<td>6.60</td>
<td>9.00</td>
</tr>
<tr>
<td>2.00</td>
<td>0.00</td>
<td>0.50</td>
<td>0.71</td>
<td>1.00</td>
<td>14.50</td>
<td>15.50</td>
</tr>
<tr>
<td>2.00</td>
<td>0.58</td>
<td>0.50</td>
<td>0.71</td>
<td>2.40</td>
<td>14.50</td>
<td>16.90</td>
</tr>
<tr>
<td>4.00</td>
<td>0.54</td>
<td>0.50</td>
<td>0.84</td>
<td>2.20</td>
<td>27.80</td>
<td>30.00</td>
</tr>
<tr>
<td>2.00</td>
<td>0.58</td>
<td>0.80</td>
<td>0.71</td>
<td>2.40</td>
<td>38.10</td>
<td>40.50</td>
</tr>
<tr>
<td>2.00</td>
<td>0.90</td>
<td>0.80</td>
<td>0.71</td>
<td>10.00</td>
<td>38.10</td>
<td>48.10</td>
</tr>
<tr>
<td>8.00</td>
<td>0.52</td>
<td>0.50</td>
<td>0.92</td>
<td>2.10</td>
<td>54.30</td>
<td>56.40</td>
</tr>
<tr>
<td>2.00</td>
<td>0.58</td>
<td>0.90</td>
<td>0.71</td>
<td>2.40</td>
<td>77.50</td>
<td>79.80</td>
</tr>
<tr>
<td>2.00</td>
<td>0.90</td>
<td>0.90</td>
<td>0.71</td>
<td>10.00</td>
<td>77.50</td>
<td>87.50</td>
</tr>
</tbody>
</table>

Table A.1: Table showing some examples of the time scales involved. The units for $\tau$, $\langle t_{\text{init}} \rangle$, $\langle t_{\text{decay}} \rangle$ and $\bar{\tau}_{1/10}$ are years. Here, the half-life $\tau$ is used to calculate the value for $s = e^{-\log(2)/\tau}$.

### A.4 Comparing this to the exponential discounting approach used by Nowak et al.

Nowak et al. behavioral model [135] assumes that the probability evolves as follows:

$$w_{t+1} = \logit^{-1} S_{t+1} = \frac{\exp S_{t+1}}{1 + \exp S_{t+1}} = \frac{1}{1 + \exp(-S_{t+1})}$$

(A.12)

where

$$S_{t+1} = v S_t + (1 - v) Y_t$$

(A.13)

where $S_t$ is the log odds of the probability $w_t$, and $Y_t$ is the terms of the change in the log odds using the EWMA framework. In the models described by Nowak et al. the term $Y_t$ can be expressed as

$$Y_t = S_0 + (1 - v)^{-1} \sum_j m_{j,t} \Delta_j,$$

(A.14)

where $j$ expresses the various levels of information and feedbacks that an agent considers in changing his/her behavior. For example these could be personal, network or media related. $\Delta_j$ is the magnitude of the change for level $j$ and $m_{j,t}$ is the weight associated to each level which in this framework can depend on time. $S_0$ is a default value that $S_t$ would tend to if the agent is no longer exposed to the updates $\Delta_j$. The interpretation of $\Delta_j$ here is a change in the log odds and as such it is not bounded in $[0, 1]$ but rather can range in $(-\infty, \infty)$.

As before in the continuous version equation A.13 can be re-expressed as $S(t) + (1 - v) S = (1 - v) Y$ which can be solved to give $S(t) = Y \left[ 1 - \exp[1 - (1 - v)t] \right]$. So if $Y_t = Y$ is a constant every year (e.g., $Y = S_0$), then in the limit of large time $S(t) = Y$ and thus $w(t) = 1/[1 + \exp(-Y)]$. 

169
A.5 Hyperbolic Discounting

As described by Farmer and Geanakoplos [54], “conventional economics supposes that agents value the present vs. the future using an exponential discounting function. In contrast, experiments with animals and humans suggest that agents are better described as hyperbolic discounters, whose discount function decays much more slowly at large times, as a power law”. A general form of hyperbolic discounting was given by Wickelgren power law, expressed as [182]

\[ m = \lambda (1 + \beta t)^{-\psi}. \]  

(A.15)

Usually, by hyperbolic discounting the power \( \psi \) is set to 1 [3, 122, 124, 131, 181]. Our model does not consider hyperbolic discounting and this will be the subject of future research.
In order to inform our microsimulation model on individual-level tax evasion and compliance behavior, we are constructing a variety of social influence parameters. In particular, we are interested in two social determinants that shape people’s attitudes towards taxes and drive their evasion / compliance. These two effects are related to people’s social networks and their exposure to media reports about taxes. While there already has been some coverage of these issues in the academic literature, we construct these measures based on data from past experiments such as the one conducted by Alm et al in 2009:

Looking at media reporting in particular, we see two potential pathways through which reports could affect taxpayer behavior: if media coverage is focused on negative aspects (such as tax fraud, inefficiency, ineffective enforcement and waste), it would cause taxpayers to lose faith in the system and evade, while, if media coverage focused on other aspects (such as stepped-up enforcement, public denunciation of famous tax evaders, public duty to pay taxes), it would deter taxpayers from evading and push them towards compliance. Consequently, we are interested in finding both the magnitude and directionality of both effects.

Similarly, when considering the effect of people’s social network, we are interested to find out how much influence these factors exert on people’s taxpaying decisions and under which circumstances these influences push people towards evasion or compliance.

In their 2009 paper from the Journal of Public Economics entitled “Getting the word out: Enforcement information dissemination and compliance behavior,” Alm and colleagues present results from a very interesting classroom experiment that we believe can be adapted to elicit the relative weights between the impacts of personal experiences, media reporting and social networks. Looking at the tiered structure and assignment into treatment conditions, it appears that it is possible to construct marginal effects as outlined in the figure below for the various social determinants of tax evasion / compliance:
Based on our understanding of the aforementioned experiment, the key treatment variables we are interested in for our margins are the constant (i.e. personal experience only), “Unofficial message allowed” (i.e. social network effect) and “Official information provided” (i.e. media reporting effect). In addition, we investigate how these margins differ by and income / wealth, age, gender, prepared on tax return, audit probability and past audit experience.

In our preliminary analyses, we specify various regression models:

\[ Y_{irs}^{i} = \beta_1 \tau_{irs} + \beta_2 X_{irs} + \gamma_{s} + \delta_{r} + \eta_{i} + \epsilon_{irs} \]

\( Y_{irs} \) is the key outcome variable, representing % compliance rate for individual \( i \) in session \( s \) and round \( r \);
\( \tau_{irs} \) is the assigned treatment for individual \( i \) in session \( s \) and round \( r \), where an individual is either left only to his or her personal judgment in deciding to comply (baseline), receives official audit information in the form of a media treatment (\( \tau_{irs} = 1 \)) or is given the opportunity to speak with his or her peers.
about taxes (τ_{34} = 2). X_{i,sr} is a set of observable individual characteristics (e.g. age, gender, income, audit probability and history of audit and penalty) for individual i. The specification includes session-fixed effects, γ_s, round-fixed effects, δ_r, and an individual-level random effects parameter η_i that accounts for covariance across estimates over time. Finally, ε_{isr} is the error term and we apply a linear panel regression framework.

### Variables

<table>
<thead>
<tr>
<th>Variables</th>
<th>Coef</th>
<th>SE</th>
<th>P-Value</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-0.4587</td>
<td>0.1256</td>
<td>0.0003</td>
<td>-0.7049 -0.2125</td>
</tr>
<tr>
<td>Age</td>
<td>0.0065</td>
<td>0.0091</td>
<td>0.4759</td>
<td>-0.0113 0.0242</td>
</tr>
<tr>
<td>Age-squared</td>
<td>0.0000</td>
<td>0.0002</td>
<td>0.8137</td>
<td>-0.0003 0.0004</td>
</tr>
<tr>
<td>Gender</td>
<td>-0.1003</td>
<td>0.0101</td>
<td>0.0000</td>
<td>-0.1201 -0.0804</td>
</tr>
<tr>
<td>Taxes prepared by a third party</td>
<td>-0.0280</td>
<td>0.0106</td>
<td>0.0083</td>
<td>-0.0487 -0.0072</td>
</tr>
<tr>
<td>Penalized</td>
<td>-0.0404</td>
<td>0.0011</td>
<td>0.0000</td>
<td>-0.0426 -0.0382</td>
</tr>
<tr>
<td>Audited</td>
<td>0.2119</td>
<td>0.0104</td>
<td>0.0000</td>
<td>0.1915 0.2322</td>
</tr>
</tbody>
</table>

### Treatment (baseline = personal info only)

| Official / media info                                | -0.1031| 0.0339  | 0.0024  | -0.1695 -0.0366 |
| Unofficial / social network info                    | 0.1131 | 0.0359  | 0.0016  | 0.0428 0.1835 |
| Penalized in previous round r-1                    | -0.0103| 0.0011  | 0.0000  | -0.0125 -0.0081 |
| Audited in previous round r-1                      | 0.0630 | 0.0104  | 0.0000  | 0.0427 0.0833 |

### Earnings in present round (baseline = $5,600)

| $6,300                                              | 0.0260 | 0.0115  | 0.0240  | 0.0034 0.0485 |
| $7,200                                              | 0.0456 | 0.0116  | 0.0001  | 0.0229 0.0683 |
| $8,100                                              | 0.0384 | 0.0119  | 0.0012  | 0.0152 0.0617 |
| $9,000                                              | 0.0241 | 0.0134  | 0.0732  | -0.0023 0.0504 |

### Wealth / LT earnings (baseline = Less than $20,000)

| $20,000 - $49,999                                   | -0.2673| 0.0315  | 0.0000  | -0.2937 -0.2409 |
| $50,000 - $74,999                                   | 0.7409 | 0.0306  | 0.0000  | 0.6809 0.8008 |
| $75,000 - $99,999                                   | 0.4547 | 0.0223  | 0.0000  | 0.4111 0.4983 |
| $100,000 - $149,999                                 | 0.2071 | 0.0147  | 0.0000  | 0.1783 0.2358 |
| $150,000 or more                                    | 0.9974 | 0.0428  | 0.0000  | 0.9134 1.0814 |

### Individual audit probability (baseline PR(audit)=0)

| Pr(audit)=0.05                                      | 0.1203 | 0.0154  | 0.0000  | 0.0901 0.1504 |
| Pr(audit)=0.10                                      | 0.1560 | 0.0146  | 0.0000  | 0.1273 0.1846 |
| Pr(audit)=0.40                                      | 0.1929 | 0.0106  | 0.0000  | 0.1721 0.2137 |

Errors clustered on individuals, controlling for state and year FE; Log-linear panel regression coefficients reported as % changes
In a variation of our regression model, we consider a more nuanced set of treatment variables, not just assessing the effect of the ability to speak to one’s network about taxes, but also the content of the specific messages sent between study participants:

\[ Y_{ist}^2 = \beta_1 \pi_{ist} + \beta_2 \mu_{ist} + \beta_3 X_{ist} + \gamma_s + \delta_r + \eta_i + \epsilon_{ist} \]

In this specification, \( \pi_{ist} \) denotes the official or media treatment and \( \mu_{ist} \) represents a vector of possible messages exchanged between individual i and other participants in a given session s and round r as part of the social network treatment (0 = "No post-audit communication allowed”; 1 = "Don’t send message”; 2 = "Not audited”; 3 = "Audited”; 4 = "Not audited + evaded”; 5 = "Not audited + complied”; 6 = "Audited + evaded”; 7 = "Audited + complied").

Since the original data from Alm and colleagues does not supply this information directly, we generated a variable, counting all the responses received by an individual respondent in each round of the experiment as part of the unofficial treatment. To control for the fact that the treatment groups have different sizes, we normalize our independent variables as shares of responses received, i.e. \( \text{Pr} (\text{Response}_s | \mu = 1) \).

<table>
<thead>
<tr>
<th>Variables</th>
<th>Model 2: Compliance Rate (log-linear panel)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coef</td>
</tr>
<tr>
<td>Constant</td>
<td>-0.5662</td>
</tr>
<tr>
<td>Age</td>
<td>0.0063</td>
</tr>
<tr>
<td>Age-squared</td>
<td>0.0000</td>
</tr>
<tr>
<td>Gender</td>
<td>-0.0995</td>
</tr>
<tr>
<td>Taxes prepared by a third party</td>
<td>-0.0281</td>
</tr>
<tr>
<td>Penalized</td>
<td>-0.0406</td>
</tr>
<tr>
<td>Audited</td>
<td>0.2131</td>
</tr>
<tr>
<td>Official / media info</td>
<td>0.1062</td>
</tr>
<tr>
<td>% of alters responding “Don’t send a message from me”</td>
<td>-0.1028</td>
</tr>
<tr>
<td>% of alters responding “I was not audited”</td>
<td>-0.1098</td>
</tr>
<tr>
<td>% of alters responding “I was audited”</td>
<td>0.0085</td>
</tr>
<tr>
<td>% of alters responding “I was not audited and I did not report all my income”</td>
<td>-0.0658</td>
</tr>
<tr>
<td>% of alters responding “I was audited and I did report all my income”</td>
<td>0.0395</td>
</tr>
<tr>
<td>% of alters responding “I was audited and I did not report all my income”</td>
<td>0.1403</td>
</tr>
<tr>
<td>Penalized in previous round r-1</td>
<td>-0.0105</td>
</tr>
<tr>
<td>Audited in previous round r-1</td>
<td>0.0638</td>
</tr>
<tr>
<td>Earnings in present round (baseline = $5,600)</td>
<td>0.0260</td>
</tr>
<tr>
<td>$6,300</td>
<td>0.0459</td>
</tr>
<tr>
<td>$7,200</td>
<td>0.0392</td>
</tr>
<tr>
<td>$8,100</td>
<td>0.0248</td>
</tr>
<tr>
<td>$9,000</td>
<td>0.0000</td>
</tr>
<tr>
<td>Wealth / LT earnings (baseline = Less than $20,000)</td>
<td>0.2696</td>
</tr>
<tr>
<td>$20,000 - $49,999</td>
<td>0.7431</td>
</tr>
<tr>
<td>$50,000 - $74,999</td>
<td>0.4558</td>
</tr>
<tr>
<td>$75,000 - $99,999</td>
<td>0.0000</td>
</tr>
</tbody>
</table>
When looking at these preliminary results, it is encouraging that the effects appear to have the anticipated direction, i.e. if alters mostly reported getting audited, individuals responded with higher rates of compliance and if alters mostly reported not getting audited or even getting away with evading their taxes, individuals were compelled to comply less. Nevertheless, perhaps due to small numbers in each category, we need to caution that none of the individual responses received show a statistically significant effect on compliance rates at conventional levels.

References

C Program Evaluation and Review Technique (PERT) distribution

The beta-PERT distribution is a smooth version of the triangular distribution that is typically used in sampling values of model parameters based on a knowledge of the minimum and maximum and an estimate of the modal values of the parameters. As described in [https://www.riskamp.com/beta-pert](https://www.riskamp.com/beta-pert), “sampling from the beta distribution requires minimum and maximum values \((x_{\text{min}} \text{ and } x_{\text{max}})\) and two shape parameters, \(v\) and \(w\). The beta-PERT distribution uses the mode or most likely parameter \((x_{\text{mode}})\) to generate the shape parameters \(v\) and \(w\) of a beta distribution. An additional scale parameter \(\lambda\) scales the height of the distribution; the default value for this parameter is 4.” In the PERT distribution, the mean \(\mu\) is calculated

\[
\mu = \frac{x_{\text{min}} + x_{\text{max}} + \lambda x_{\text{mode}}}{\lambda + 2}, \quad \text{(C.1)}
\]

and is used to calculate the \(v\) and \(w\) shape parameters

\[
v = \frac{(\mu - x_{\text{min}})(2x_{\text{mode}} - x_{\text{min}} - x_{\text{max}})}{(x_{\text{mode}} - \mu)(x_{\text{max}} - x_{\text{min}})}, \quad \text{(C.2)}
\]

\[
w = \frac{(x_{\text{max}} - \mu)(2x_{\text{mode}} - x_{\text{min}} - x_{\text{max}})}{(x_{\text{mode}} - \mu)(x_{\text{max}} - x_{\text{min}})}. \quad \text{(C.3)}
\]
D  Survey Questions of the American Life Panel Well Being 456

In this appendix, we list the questions asked in our ALP survey. However, before we describe and list the survey questions we first describe the available socioeconomic and demographic variables of the respondents and the questions used by the ALP surveys.

D.1  Socio-Economic and Demographic Variables

Table D.1 and D.2 shows the questions regarding socio-economic and demographic variables. Of particular relevance to this project are gender, age, family income, do they work, highest education and whether we can identify non-US born respondents that moved to the US during adulthood.

<table>
<thead>
<tr>
<th>Question ID</th>
<th>Question Text</th>
</tr>
</thead>
<tbody>
<tr>
<td>gender</td>
<td>What is your gender?</td>
</tr>
<tr>
<td>calcage</td>
<td>What is your age?</td>
</tr>
<tr>
<td>birthyear</td>
<td>Year</td>
</tr>
<tr>
<td>currentlivingsituation</td>
<td>Could you tell us what your current living situation is? 1 Married or living with a partner; 2 Separated; 3 Divorced; 4 Widowed; 5 Never married</td>
</tr>
<tr>
<td>borninus</td>
<td>Were you born in the United States?</td>
</tr>
<tr>
<td>stateborn</td>
<td>In what state were you born?</td>
</tr>
<tr>
<td>citizenus</td>
<td>Are you a citizen of the United States?</td>
</tr>
<tr>
<td>familyincome</td>
<td>Which category represents the total combined income of all members of your family (living here) during the past 12 months? This includes money from jobs, net income from business, farm or rent, pensions, dividends, interest, social security payments and any other money income received by members of your family who are 15 years of age or older. 1) Less than $5,000; 2) $5,000 to $7,499; 3) $7,500 to $9,999; 4) $10,000 to $12,499; 5) $12,500 to $14,999; 6) $15,000 to $19,999; 7) $20,000 to $24,999; 8) $25,000 to $29,999; 9) $30,000 to $34,999; 10) $35,000 to $39,999; 11) $40,000 to $49,999; 12) $50,000 to $59,999; 13) $60,000 to $74,999; 14) $75,000 or more</td>
</tr>
<tr>
<td>familyincome_part2</td>
<td>You told us that the total combined income of all members of your family (living here) during the preceding 12 months was more than $75,000. Thinking about the total combined income of your family from all sources, approximately how much did members of your family receive during the previous 12 months? 1) $75,000-$99,999; 2) $100,000-$124,999; 3) $125,000-$199,999; 4) $200,000 or more</td>
</tr>
</tbody>
</table>

Table D.1: Socio-economic and demographic questions.
<table>
<thead>
<tr>
<th>Question ID</th>
<th>Question Text</th>
</tr>
</thead>
<tbody>
<tr>
<td>householdmembers</td>
<td>Now we would like to know about other members of your household, if there are any. How many other people live with you? (enter 0 for no one else).</td>
</tr>
<tr>
<td>doyouwork</td>
<td>Next are some questions about your current, main job. Do you work for someone else, are you self-employed, or what? 1 Work for someone else; 2 Self-employed; 3 Other</td>
</tr>
<tr>
<td>highesteducation</td>
<td>What is the highest level of school you have completed or the highest degree you have received? 1) Less than 1st grade; 2) 1st,2nd,3rd,or 4th grade; 3) 5th or 6th grade; 4) 7th or 8th grade; 5) 9th grade; 6) 10th grade; 7) 11th grade; 8) 12th grade NO DIPLOMA; 9) HIGH SCHOOL GRADUATE high school DIPLOMA or the equivalent (For example: GED); 10) Some college but no degree; 11) Associate degree in college Occupational/vocational program; 12) Associate degree in college Academic program; 13) Bachelor’s degree (For example: BA,AB,BS); 14) Master’s degree (For example: MA,MS,MEng,MEd,MSW,MBA); 15) Professional School Degree (For example: MD,DDS,DVM,LLB,JD); 16) Doctorate degree (For example: PhD,EdD)</td>
</tr>
<tr>
<td>currentjobstatus</td>
<td>What is your current employment situation? 1 Working Now; 2 Unemployed and looking for work; 3 Temporarily laid off, on sick or other leave; 4 Disabled; 5 Retired; 6 Homemaker; 7 Other</td>
</tr>
<tr>
<td>statereside</td>
<td>Now we would like to know about where you live. In which state do you reside?</td>
</tr>
<tr>
<td>ethnicity</td>
<td>Do you consider yourself primarily white or Caucasian, Black or African American, American Indian, or Asian? 1 White/Caucasian; 2 Black/African American; 3 American Indian or Alaskan Native; 4 Asian or Pacific Islander; 5 Other</td>
</tr>
<tr>
<td>ethnicity_pacificislander</td>
<td>Are you Asian or Pacific Islander?</td>
</tr>
<tr>
<td>hispaniclatino</td>
<td>Do you consider yourself Hispanic or Latino?</td>
</tr>
<tr>
<td>hispaniclatino_detail</td>
<td>Would you say that you are primarily Mexican American, Puerto Rican, Cuban, or something else? 1 Mexican American; 2 Puerto Rican; 3 Cuban; 4 Something else</td>
</tr>
<tr>
<td>tshhbox</td>
<td>Time at which demographic variables were collected</td>
</tr>
<tr>
<td>recruitment_type</td>
<td>Type of recruitment</td>
</tr>
<tr>
<td>webtv</td>
<td>Hardware provided to respondent</td>
</tr>
<tr>
<td>oldprim_key</td>
<td>Previous primary key.</td>
</tr>
</tbody>
</table>

Table D.2: Socio-economic and demographic questions.
D.2 Survey Questions of the American Life Panel Well Being 456

The introduction of the survey that was presented to the respondents read. At the beginning of the survey, we associated four random variables to each respondent which determine which version of certain questions they would be asked. These random variables were

- *AuditRandom* ranging in values of 1, 2 and 3.
- *BehavReactionRandom* ranging in values of 1 and 2.
- *BehavReactionRandom2* ranging in values of 1 and 2.
- *BehavReactionRandom3* ranging in values of 1 and 2.

Table D.3 shows questions that asked about the people they know and regularly interact with.

<table>
<thead>
<tr>
<th>Qu.#</th>
<th>Question ID</th>
<th>Question Text</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>alters</td>
<td>Please list the initials of 10 adults that you know, other than spouses or domestic partners, and interact with on a regular basis.</td>
</tr>
<tr>
<td>2</td>
<td>alterrel</td>
<td>For each of the people listed below, please indicate your primary relationship with that person. a. Family member b. Friend c. Coworker d. Other.</td>
</tr>
<tr>
<td>3</td>
<td>altereduc</td>
<td>For each of the people listed below, please indicate, to the best of your knowledge, what is the highest education degree that person has received? a. Less than a high school diploma or the equivalent (For example: GED) b. High school diploma or the equivalent (For example: GED) c. Associate degree in college d. Bachelor’s degree (For example: BA,AB,BS) e. Graduate degree, such as Master’s or Doctoral-level degree</td>
</tr>
<tr>
<td>4</td>
<td>altertalktax</td>
<td>For each of the people listed below, please check the box next to any person with whom you have talked or consulted with about taxes in the past 5 years. This could include any aspect of taxes, including state or federal taxes, tax audits or penalties, how fair taxes seem, or any other related topic. a. Yes b. No c. Don’t know d. I would prefer not to say</td>
</tr>
<tr>
<td>5</td>
<td>taxhowoften</td>
<td>For these people, how often do you talk to them about taxes? a. Once every five years b. Once every two years c. Once a year d. Twice a year e. Monthly f. More frequently g. I don’t know or don’t remember</td>
</tr>
<tr>
<td>6</td>
<td>taxselfemployed</td>
<td>For each of the people below, do you think they are self-employed or have rental income? a. Yes b. No c. I don’t know or don’t remember</td>
</tr>
<tr>
<td>7</td>
<td>taxaudit</td>
<td>For each of the people below, do you know or think that they have been audited by the IRS in the past five years? a. Yes, I know they have been audited b. Yes, I think they have been audited c. I don’t know or don’t remember d. No, I don’t think they have been audited e. No, I know they have not been audited.</td>
</tr>
</tbody>
</table>

Table D.3: Survey questions on the respondent’s social network.

Table D.4 shows the questions asked in Section 3 to 6 of the ALP survey. These questions ask about the respondent’s thoughts and experiences regarding US federal income taxes. Specifically, they ask you about three aspects of how income taxes work:

1. The audit rate: The percentage of taxpayers whose returns are audited by the IRS.
2. The penalty rate: The size of the penalty for not paying all of your owed taxes, and
3. The effective income tax rate: The percentage of your income that you owe to the government in taxes.

<table>
<thead>
<tr>
<th>Qu.#</th>
<th>Question ID</th>
<th>Question Text</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>perceivedauditrate &amp; -magnifier</td>
<td>In a typical year, what percent of taxpayers in the U.S. will have their income tax return audited by the IRS?</td>
</tr>
<tr>
<td>9a</td>
<td>perceivedaru - underreport</td>
<td>Imagine a taxpayer that only paid [30% of the taxes he or she owes, 60% of the taxes he or she owes, 90% of the taxes he or she owes]. Do you think the chances of that person being audited that year would be higher, lower, or the same as if he or she had paid all taxes owed?</td>
</tr>
<tr>
<td>9b</td>
<td>perceivedaruprob &amp; -magnifier</td>
<td>What is the percent chance that person will have their income tax return audited by the IRS?</td>
</tr>
<tr>
<td>10a</td>
<td>bombardcrater</td>
<td>If your tax return was audited last year, do you think the chances of being audited the following year are higher, lower, or the same?</td>
</tr>
<tr>
<td>10b</td>
<td>bombardcrateramount &amp; -magnifier</td>
<td>What do you think are these new chances of getting audited?</td>
</tr>
<tr>
<td>11</td>
<td>perceivedpenaltyrate</td>
<td>Now let’s consider the penalty rate. If the IRS detects that a person has underreported their taxes, they will first have to pay the unpaid taxes that were due. In addition, they will be assessed a penalty that is a percentage of the amount they underpaid. This percentage is the penalty rate. Imagine a person was caught underpaying their taxes by $1000. In addition to having to pay that $1000, how much of a penalty would they have to pay? <strong>NOTE:</strong> We converted this response to a proportion so that in our analysis perceivedpenaltyrate represents a proportion relative to $1000.</td>
</tr>
<tr>
<td>12</td>
<td>perceivedtaxrate</td>
<td>What do you think your effective income tax rate was this past year? Visual 0-100 subjective probability slider [with checks that response is between 0 and 100]</td>
</tr>
<tr>
<td>13</td>
<td>perceivedevasionratepopulation</td>
<td>In a typical year, out of all taxpayers in the United States, what percent intentionally underreport their taxes?</td>
</tr>
<tr>
<td>14</td>
<td>perceivedevasionrate</td>
<td>Now consider people like you. In a typical year, out of 100 people like you, how many intentionally underreport their taxes?</td>
</tr>
<tr>
<td>15</td>
<td>perceivedevasion-manyevaders</td>
<td>Imagine that a widely-disseminated news story comes out that half of all US taxpayers underreport their taxes. Out of 100 people like you, how many would now underreport their taxes?</td>
</tr>
<tr>
<td>16</td>
<td>perceivedcaught</td>
<td>In a typical year, what percent will be caught by the IRS?</td>
</tr>
</tbody>
</table>

Table D.4: Notice that some questions required a Magnifier scale as explained in the text.

When asked to estimate the risk of being audited as a percentage, respondents were presented with a slider bar allowing them to enter (or directly type) the value as an integer number. Since, many respondents may correctly believe that this risk is smaller than 1%, for those who gave either 0% or 1% on the percentage slider bar we asked a Magnifier question which read:
We would like to get extra information about the last question. What do you think are these new chances of getting audited? 0%, More than 0% and less than or equal to .001% (1/100,000), More than .001% (1/100,000) and less than or equal to .01% (1/10,000), More than .01% (1/10,000) and less than or equal to .1% (1/1,000), More than .1% (1/1,000) and less than or equal to 1% (1/100), 1%.

We used the responses to the magnifier questions to replace and better inform the perceived audit risks of respondents for the various questions regarding the risk of being audited.

Tables D.5 and D.6 show the questions asked in Section 7 of the ALP survey and considered the perceived behavioral reaction to increased or decreased tax rate. Provided that the respondent gave a non-zero answer for perceived evasion rate, then based on the previously determined random variable BehavReactionRandom, respondents were randomly presented the question in either of the two tables. The introduction to these questions read:

Now let’s consider the effect of changing the effective income tax rate (but the audit and penalty rates remain unchanged). As a reminder, you stated earlier that [PerceivedEvasionRate] out of 100 of people like you underreport their taxes to the IRS.

<table>
<thead>
<tr>
<th>Qu.#</th>
<th>Question ID</th>
<th>Question Text</th>
</tr>
</thead>
<tbody>
<tr>
<td>17</td>
<td>perveivedunderreport-taxhigher</td>
<td>Imagine instead that people’s effective income tax rates were 50% higher than they currently are. Out of 100 US taxpayers like you, who continue to work just as many hours, how many do you think would intentionally underreport their taxes to the IRS?</td>
</tr>
<tr>
<td>18</td>
<td>perveivedunderreport-taxmuchhigher</td>
<td>Imagine instead that people’s effective income tax rates were twice as high as they currently are. Out of 100 US taxpayers like you, who continue to work just as many hours, how many do you think would intentionally underreport their taxes to the IRS?</td>
</tr>
</tbody>
</table>

Table D.5

<table>
<thead>
<tr>
<th>Qu.#</th>
<th>Question ID</th>
<th>Question Text</th>
</tr>
</thead>
<tbody>
<tr>
<td>19</td>
<td>perveivedunderreport-taxlower</td>
<td>Imagine instead that people’s effective income tax rates were 25% lower than what they currently are. Out of 100 US taxpayers like you, who continue to work just as many hours, how many do you think would intentionally report less income to the IRS than they actually earned?</td>
</tr>
<tr>
<td>20</td>
<td>perveivedunderreport-taxmuchlower</td>
<td>Imagine instead that people’s effective income tax rates were half of what they currently are. Out of 100 US taxpayers like you, who continue to work just as many hours, how many do you think would intentionally report less income to the IRS than they actually earned?</td>
</tr>
</tbody>
</table>

Table D.6
Tables D.7 show the questions asked in Section 8 of the ALP survey and considered the perceived behavioral reaction to increased audit rate.

<table>
<thead>
<tr>
<th>Qu.#</th>
<th>Question ID</th>
<th>Question Text</th>
</tr>
</thead>
<tbody>
<tr>
<td>21</td>
<td>perceivedunderreport-audithigher</td>
<td>Imagine that the audit rate was twice as high as it currently is. Out of 100 US taxpayers like you, how many do you think would intentionally underreport their taxes to the IRS?</td>
</tr>
<tr>
<td>22</td>
<td>perceivedunderreport-auditmuchhigher</td>
<td>Imagine that the audit rate was three times as high as it currently is. Out of 100 US taxpayers like you, how many do you think would intentionally underreport their taxes to the IRS?</td>
</tr>
</tbody>
</table>

Table D.7

Tables D.8 and D.9 show the questions asked in Section 9 of the ALP survey and considered the perceived behavioral reaction to increase or decreases of the penalty rate. Provided that the respondent gave a non-zero answer for perceivedevasionrate, then based on the previously determined random variable BehavReactionRandom, respondents were randomly presented the question in either of the two tables. The introduction to these questions read:

Now let’s consider the effect of changing the penalty rate (but the tax and audit rates remain unchanged). If the federal government detects that you have underreported how much taxes you owe, you will have to pay the unpaid taxes that were due. In addition, you will be assessed a penalty that is a percentage of the amount of taxes due that were unpaid. This percentage is the penalty rate.

<table>
<thead>
<tr>
<th>Qu.#</th>
<th>Question ID</th>
<th>Question Text</th>
</tr>
</thead>
<tbody>
<tr>
<td>23</td>
<td>perceivedunderreport-penaltyhigher</td>
<td>Imagine instead that the penalty rate was 50% higher than it currently is. Out of 100 US taxpayers like you, how many do you think would intentionally underreport their taxes to the IRS?</td>
</tr>
<tr>
<td>24</td>
<td>perceivedunderreport-penaltymuchhigher</td>
<td>Now imagine that the penalty rate was twice as high as it currently is. Out of 100 US taxpayers like you, how many do you think would intentionally underreport their taxes to the IRS?</td>
</tr>
</tbody>
</table>

Table D.8

Table D.11 show the questions relating to the Perceived behavioral reaction past refund/tax debt.

Question 28 in the survey provides also considers a hypothetical case. There were two versions of this question: one where respondents had to consider receiving a $1000 refund in the previous year, and the other where they owed an additional $1000 in the previous year to the IRS. We call these the refund and the debt groups respectively. Respondents were split roughly equally into the two groups.
<table>
<thead>
<tr>
<th>Qu.#</th>
<th>Question ID</th>
<th>Question Text</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>perceivedunderreport-penaltymuchlower</td>
<td>Imagine instead that the penalty rate was half of what it currently is. Out of 100 US taxpayers like you, how many do you think would intentionally underreport their taxes to the IRS?</td>
</tr>
<tr>
<td>26</td>
<td>perceivedunderreport-penaltylower</td>
<td>Imagine instead that the penalty rate was 25% lower than it currently is. Out of 100 US taxpayers like you, how many do you think would intentionally underreport their taxes to the IRS?</td>
</tr>
</tbody>
</table>

Table D.9

<table>
<thead>
<tr>
<th>Qu.#</th>
<th>Question ID</th>
<th>Question Text</th>
</tr>
</thead>
<tbody>
<tr>
<td>27</td>
<td>perceivedunderreport-auditpenalty_([a,b,...,h])</td>
<td>Let’s consider how low the effective income tax rate would need to be before everyone reported 100% of their taxes to the IRS, assuming there are no audits or penalties. For this question, assume for the moment that everyone has the same effective tax rate. If each of the effective income tax rates below were applied to everyone, please indicate if you think (a) the majority of people like you would report their full income OR (b) the majority of people like you would underreport their income: (i) Income tax rate = 1% (ii) Income tax rate = 2.5% (iii) Income tax rate = 5% (iv) Income tax rate = 10% (v) Income tax rate = 15% (vi) Income tax rate = 20% (vii) Income tax rate = 25% (viii) Income tax rate = 30%</td>
</tr>
</tbody>
</table>

Table D.10: Survey question asking respondents about the tax rate at which taxpayers like themselves would comply to even if their perceived risk of being audited is negligible. Note however, this question has a problem as it starts off by asking the respondents to consider how low the effective income tax rate would need to be before everyone reported 100% of their taxes to the IRS, assuming there are no audits or penalties, but then as for the rate where the majority of people would be compliant.

<table>
<thead>
<tr>
<th>Qu.#</th>
<th>Question ID</th>
<th>Question Text</th>
</tr>
</thead>
<tbody>
<tr>
<td>28</td>
<td>refunddebt</td>
<td>Imagine for the moment that last year you [received a $1000 refund/owed an additional $1000] on your federal income taxes. You are now preparing your taxes for this year. You are considering claiming a $1000 deduction, but to the best of your own knowledge you are not entirely sure if it is appropriate for you to take. Without consulting anyone else, what is the percent chance that you will claim this deduction?</td>
</tr>
</tbody>
</table>

Table D.11

Table D.12 show the questions relating tax fairness-related considerations and the importance of personal, network, and media information on fairness and on audit/penalty risk.

Table D.13 relating to the perception of services provided by taxes and the effect of learning that a famous actor was caught and prosecuted for tax evasion.

Table D.14 show the questions relating to the perception of free-riding behavior.
<table>
<thead>
<tr>
<th>Qu.#</th>
<th>Question ID</th>
<th>Question Text</th>
</tr>
</thead>
</table>
| 29   | taxesimportant-\_\{a,b,c,d\} | Please allocate 100 tokens to the issues below. More tokens means more important. In terms of how you think about taxes and paying your taxes, how important is each of the following?  
   a. The amount of taxes that I owe (that is, your effective tax rate)  
   b. The cost (in time and money) to figure out my taxes  
   c. Benefits and public services supported by taxes (for example, public education, security, welfare programs)  
   d. A moral obligation to correctly report and pay all my taxes |
| 30   | taxesfairness\_\{a,b,c\} | Now let’s consider your thoughts on the fairness of taxes, and what you’ve seen and heard from those around you. Again, please allocate 100 tokens to the issues below. In terms of how fair taxes seem to you, how important is each of the following?  
   a. Your own thoughts on the fairness of taxes and the tax system (for example, equity in how different people are taxed and how tax revenue is used by the government)  
   b. What you hear and know from friends, family, and other close contacts about the fairness of taxes and the tax system (for example, how often people cheat on their taxes)  
   c. What you hear broadly from the media and other sources about the fairness of taxes and the tax system (for example, how often people cheat on their taxes) |
| 31   | riskauditspenalties-\_\{a,b,c\} | Now let’s consider your thoughts on the risk of audits and penalties for not paying one’s taxes. Again, please allocate 100 tokens to the issues below. In terms of how you think about these risks, how important is each of the following?  
   a. Your own thoughts on the risk of audits and penalties if you don’t pay your taxes  
   b. What you hear and know from friends, family, and other close contacts about audits and penalties for not paying taxes  
   c. What you hear broadly from the media and other sources about audits and penalties for not paying taxes |

Table D.12
<table>
<thead>
<tr>
<th>Qu.#</th>
<th>Question ID</th>
<th>Question Text</th>
</tr>
</thead>
<tbody>
<tr>
<td>32</td>
<td>servicestaxes</td>
<td>Many public goods and services, such as interstate highways, national defense, national parks, and environmental protection, are in part paid for by federal income taxes. To what extent are the public goods and services that you receive worth the federal income taxes you pay? 1 = Not at all worth it, 2, 3, 4, 5 = Definitely worth it</td>
</tr>
<tr>
<td>43</td>
<td>actor</td>
<td>Imagine that you heard a famous actor was caught and prosecuted for tax evasion. In your mind, would hearing about this make you more or less likely to report all of taxes you owe to the IRS? 1 = I would be much more likely to fully report my income, 2, 3, 4, 5 = I would be much less likely to fully report my income. <strong>NOTE:</strong> In our analyses we decided to invert the scale so that 1 = I would be much less likely to fully report my income and 5 = I would be much more likely to fully report my income.</td>
</tr>
</tbody>
</table>

Table D.13

<table>
<thead>
<tr>
<th>Qu.#</th>
<th>Question ID</th>
<th>Question Text</th>
</tr>
</thead>
<tbody>
<tr>
<td>33-37</td>
<td>fr1_{1,2,...,5}</td>
<td>Imagine each of the following situations. For each, please indicate if it is always OK to engage in the described behavior, sometimes OK to do it, or never OK to do that behavior. 1. Regularly listen to public radio without ever contributing. 2. Illegally copying, downloading, or streaming movies. 3. Have a dog but not getting it spayed or neutered. 4. Avoid getting the flu vaccine. 5. Avoid paying all of the income tax that you owe.</td>
</tr>
<tr>
<td>38-42</td>
<td>fr2_{1,2,...,5}</td>
<td>Imagine each of the following situations. For each, how many people out of 100 would say that it is at least sometimes OK to engage in the described behavior 1. Regularly listen to public radio without ever contributing. 2. Illegally copying, downloading, or streaming movies. 3. Have a dog but not getting it spayed or neutered. 4. Avoid getting the flu vaccine. 5. Avoid paying all of the income tax that you owe.</td>
</tr>
</tbody>
</table>

Table D.14
<table>
<thead>
<tr>
<th>Qu.#</th>
<th>Question ID</th>
<th>Question Text</th>
</tr>
</thead>
<tbody>
<tr>
<td>44</td>
<td>workforpay</td>
<td>Have you ever worked for pay?</td>
</tr>
<tr>
<td>45</td>
<td>selfemployed</td>
<td>Do you work for someone else, are you self-employed, or what? Work for someone else; Self-employed; Other.</td>
</tr>
<tr>
<td>46</td>
<td>everfiledtaxes</td>
<td>Have you ever filed a tax return yourself or had someone file it for you? Yes; No; Don’t know or can’t remember.</td>
</tr>
<tr>
<td>47</td>
<td>preptaxes</td>
<td>Do you typically prepare your own tax returns or do you pay someone (e.g., an accountant or lawyer) to prepare them for you? I prepare my own tax returns using tax software on the computer; I prepare my own tax returns, without using tax software; I pay someone else to prepare my tax returns; I do not prepare a tax return; I don’t know or would prefer not to say.</td>
</tr>
<tr>
<td>48</td>
<td>everaudited</td>
<td>Have you ever been audited by the IRS? Yes; No; Don’t know or can’t remember; Prefer not to answer.</td>
</tr>
<tr>
<td>49</td>
<td>spouseaudit</td>
<td>Has your spouse or domestic partner ever been audited by the IRS at any time during the past five years? I am not currently married or living with a domestic partner; Yes; No; Don’t know or can’t remember; Prefer not to answer.</td>
</tr>
</tbody>
</table>

Table D.15: Survey questions on employment and on federal income tax preparation and audit experiences.
E  Fitting data to an S-Curve

Let’s assume we have a data set with two fields for variables $x$ and $y$ and both are continuous and can range in the $[0, 1)$ interval. Let’s further assume that we believe that $y$ depends on $x$ and varies according to some sort of S-shaped or sigmoid curve such that for $x = 0$ we have $y = 0$. We would like to find an S-shaped function that can best fit the data.

E.1 Fitting data to a log-normal cumulative distribution S-Curve

If we use assume that our S-shape follow the functional form of CDF of a log normal distribution we can express this dependence as

$$y(x, m, s) = \frac{1}{2} + \frac{1}{2} \text{erf} \left( s \cdot \frac{(\ln x - \ln m)}{\sqrt{2}} \right), \quad (E.1)$$

where erf is the error function. When $x$ is equal to $m$ the value of $y$ is 50%. The parameter $s$ is related to how steep the curve becomes in the neighborhood of $x = m$. Specifically, the gradient of the curve at $x = m$ can be found from the functional form of the probability distribution function of the log-normal distribution. It is given by

$$g_m = \frac{s}{m\sqrt{2\pi}}. \quad (E.2)$$

Therefore, $s$ controls the steepness of the curve. For each row or data entry $i$, let’s denote the data for columns $x$ and $y$ as $x^{(i)}$ and $y^{(i)}$. By defining $Y^{(i)} = \text{inv.erf} \left( 2y^{(i)} - 1 \right)$ and $X^{(i)} = \ln x^{(i)}$, equation [E.1] can be expressed as $Y^{(i)} = kX^{(i)} + c$. Here, inv.erf denotes the inverse error function. Therefore, by creating the transformed dataset for $Y^{(i)}$ and $X^{(i)}$ and carrying out a linear regression of $Y$ on $X$ we can estimate the values for $k$ and $c$. The best fit parameters for $m$ and $s$ found from $k$ and $c$ as follows: $s = k\sqrt{2}$ and $m = e^{-c/k}$.

E.2 Fitting data to a logistic S-Curve function

If we use assume that our S-shape follow the logistic functional form we can express this dependence as

$$y(x, m, g_m) = \{1 + \exp[-4g_m(x - m)]\}^{-1}, \quad (E.3)$$

where $m$ represents the value of $x$ where $y$ is 50% and $g_m$ is the gradient of the S-Curve as $x = m$. Using data that provides the $y$ values for various $x$ values we can estimate the parameters $m$ and $g_m$ by a logistic regression as expressed by

$$\text{logit}[y(x, m, g_m)] = \log \left[ \frac{y(x, m, g_m)}{1 - y(x, m, g_m)} \right] = -4g_m(m - x) = \beta_0 + \beta_x x. \quad (E.4)$$

One important limitation of the logistic S-Curve function compared to the log-normal S-Curve is that generally $y$ is positive and non-zero at $x = 0$ at $y$ is not necessarily equal to 100% when $x$ reaches 100%.
Estimating Elasticities

We used cleaned American Life Panel Tax Evasion Survey data to estimate tax evasion elasticities with respect to effective tax rate, audit rate and penalty rate. The cleaned dataset had responses from 822 respondents. Each respondent received questions about six hypothetical scenarios and one baseline scenario. In the baseline scenario, we asked them about perceived tax evasion rate among 100 people like them, their perceived effective tax rate, perceived audit rate and perceived penalty for evading $1000 in taxes. In the hypothetical scenarios, we asked what would be the evasion rate under different values of tax rate, audit rate and penalty rates. All these rates (evasion, tax, audit and penalty) were transformed into a percentage scale. These variables then were included into several log-linear regression models to estimate the elasticities.

The dependent variable in all these models is logarithmic transformation of evasion rate (perceived tax evasion rate among 100 people like respondent), \(\ln E\). The main independent variables of interest are logarithmic transformations of tax rate \(\ln T\), audit rate \(\ln A\) and penalty rate \(\ln P\). The coefficients for these variables had the expected signs and were significant in all considered models. The absolute values of these coefficients are less than 1 in all considered models. This means that tax evasion is inelastic with respect to tax rate, audit rate and penalty rate. However, since these are log-linear models, the estimates assume constant elasticity. Therefore, at different points, the elasticities might be elastic if the constant elasticity assumption is relaxed.

Below are the STATA outputs for the selected models that we ran. Some other covariates that we used in these models were gender (= 1 if male; = 0 otherwise), natural log of family income \((lnI)\), natural log of respondent’s age \((lnAge)\), whether or not respondent prepares his/her tax return himself/herself \((preptaxeself) = 1 \text{ if yes}; = 0 \text{ otherwise})\), if respondent has ever been audited or not \((everaudited = 1 \text{ if yes}; = 0 \text{ otherwise})\), if respondent self-employed or not \((selfemployed1 = 1 \text{ if self-employed}; = 0 \text{ otherwise})\), and if respondent thinks that it does not worth paying taxes for public goods and services that he/she received \((worthayingtaxes = 1 \text{ if does not worth}; = 0 \text{ otherwise})\).
Model 1

Fixed-effects (within) regression

Number of obs = 3989
Number of groups = 732

R-sq: within = 0.4219
between = 0.0000
overall = 0.0118

F(731, 3254) = 22.05           Prob > F = 0.0000

corr(u_i, Xb) = -0.7419

|              | Coef.  | Std. Err. | t     | P>|t|   | [95% Conf. Interval] |
|--------------|--------|-----------|-------|-------|----------------------|
| lnE          |        |           |       |       |                      |
| lnT          | .9975227 | .0344718  | 28.94 | 0.000 | .929934 1.065111    |
| lnA          | -.6212663 | .0213063  | -29.16 | 0.000 | -.6630414 -.5794913 |
| lnF          | -.6362474 | .0349893  | -18.18 | 0.000 | -.7048507 -.567644  |
| _cons        | -4.585937 | .1759326  | -26.07 | 0.000 | -4.930887 -4.240987 |

sigma_u = 1.5611081
sigma_e = .51494973
rho = .9018688 (fraction of variance due to u_i)

F(3,3254) = 791.73           Prob > F = 0.0000

Model 2

Random-effects GLS regression

Number of obs = 3989
Number of groups = 732

R-sq: within = 0.4011
between = 0.0017
overall = 0.0210

Wald chi2(3) = 1415.06           Prob > chi2 = 0.0000

corr(u_i, X) = 0 (assumed)

|              | Coef.  | Std. Err. | t     | P>|t|   | [95% Conf. Interval] |
|--------------|--------|-----------|-------|-------|----------------------|
| lnE          |        |           |       |       |                      |
| lnT          | .9059645 | .0317981  | 28.49 | 0.000 | .8436413 .9682876    |
| lnA          | -.3538995 | .0181169  | -19.78 | 0.000 | -.393908 -.322891   |
| lnF          | -.276096 | .0254309  | -10.86 | 0.000 | -.3259396 -.2262524 |
| _cons        | -2.681604 | .1307805  | -19.89 | 0.000 | -3.597929 -1.855279  |

sigma_u = .91574958
sigma_e = .51494973
rho = .75975692 (fraction of variance due to u_i)
Model 3

Random-effects ML regression
Number of obs = 3972
Group variable: respid
Number of groups = 729

Random effects u_i ~ Gaussian
Obs per group: min = 1
avg = 5.4
max = 7

Log likelihood = -4317.6455
LR chi2(5) = 1239.25
Prob > chi2 = 0.0000

|     | Coef.   | Std. Err. | z     | P>|z|    | [95% Conf. Interval] |
|-----|---------|-----------|-------|--------|----------------------|
| lE  | .9640914  | .0321022  | 30.03 | 0.000  | 0.9011723 1.027011   |
| lnA | -.445292  | .0205579  | -21.66 | 0.000  | -.4855848 -.4049992 |
| lnF | -.3696378  | .0292509  | -12.64 | 0.000  | -.4269686 -.312307  |
| lnf | -.2828263  | .0577831  | -4.89  | 0.000  | -.396079 -.1695736  |
| lnAgp| -2.2461519  | .182023   | -1.35  | 0.176  | -.6029105 .1106366  |
| _cons | 1.021767  | .9243213  | 1.11   | 0.269  | -.7898692 2.833404   |

Likelihood-ratio test of sigma_u=0: chibar2(01) = 3623.93 Prob>chibar2 = 0.000

Model 4

Random-effects GLS regression
Number of obs = 2544
Group variable: respid
Number of groups = 455

R-sq: within = 0.4153
between = 0.0018
overall = 0.0255

Wald chi2(10) = 986.71
Prob > chi2 = 0.0000

corr(u_i, X) = 0 (assumed)

|     | Coef.   | Std. Err. | z     | P>|z|    | [95% Conf. Interval] |
|-----|---------|-----------|-------|--------|----------------------|
| lnE | .9677134  | .0417852  | 23.16 | 0.000  | .885816 .1.049611    |
| lnA | -.3913407  | .0234474  | -16.69 | 0.000  | -.4372968 -.3453846 |
| lnF | -.3116907  | .0331421  | -9.40  | 0.000  | -.3766481 -.2467333 |
| lnf | -.1375199  | .0631122  | -2.18  | 0.033  | -.2612175 -.0138223 |
| lnAgp| -2.657282  | .2058471  | -2.75  | 0.006  | -.969181 -.1622753 |
| gender | -.2388372  | .1001722  | -2.38  | 0.017  | -.4499711 -.0423033 |
| pretaxasself | -.2056623  | .1009886  | -2.04  | 0.041  | -.4039963 -.0081285 |
| everauditedself | -.0608042  | .1314828  | -0.46  | 0.644  | -.3185058 .1968975 |
| worthpayingtaxes | .1314393  | .1538662  | 0.86   | 0.391  | -.1696238 .433516  |
| selfemployed | .0965379  | .1071577  | 0.90   | 0.368  | -.1134873 .3065631 |
| _cons | 1.182164  | .9746401  | 1.21   | 0.225  | -.7209553 3.092424   |

sigma_u | .92073536
sigma_u | -.7533871
rho | .75535871  (fraction of variance due to u_i)
S-Curve Relationship between Evasion Rate and Tax Rate

Several other models were considered too. These models assumed sigmoid or S-curve relationship between evasion rate and effective tax rate. In one of these models the functional form of this relationship was a log normal distribution:

\[
E^{(l)} = \frac{1}{2} + \frac{1}{2} \text{erf} \left[ s \cdot \frac{(lnT^{(l)} - ln m)}{\sqrt{2}} \right]
\]

The parameters \( s, m \) and \( g \) were estimated based on this model. Before estimating these parameters, we imposed certain conditions in the dataset. These conditions are the following:

1) Eighth scenario was added where tax rate \( T \) was equated to 0.7 and the evasion rate \( E \) was set to 0.99;
2) If \( E = 0 \), then it was set to be 0.0001, a really small but non-zero value.
3) If \( E > 0.98 \), then it was set to be 0.98, except for \( E \) in the eighth scenario;
4) If \( T > 0.7 \), then it was replaced with missing value;
5) If \( T = 0 \), then its logarithmic transformation was set to be \( ln(0.0001) \), i.e. \( ln(T) = ln(0.0001) \)

It should be mentioned that the estimates for the parameters \( s, m \) and \( g \) were very sensitive to the conditions described above.

As an alternative to the model above, it was assumed that the S-curve relationship follows logit distribution. The equation for this model is:

\[
\ln \left( \frac{E}{1-E} \right) = \beta_0 + \beta_1 T + \beta_2 A + \beta_3 P + \beta_4 Income + \beta_5 Gender + \beta_6 Age,
\]

where \( E \) - evasion rate, \( T \) - tax rate, \( A \) - audit rate, \( P \) - penalty rate, \( Income \) - family income in $, \( Gender = 1 \) if respondent is male and 0 otherwise, \( Age \) - respondent’s age. STATA output for the estimated logit model is given below:
Logit-based S-Curve Model

At the mean values of A ( = 0.181), P ( = 0.033), income ( = $83,329.28), and Age ( = 57.7) variables, the logit model yielded $m = 0.37634$ for males and slightly higher, $0.38547$ for females. The $g$, slope of the curve at the point of $T = m$, was estimated to be 2.7923075. The following mathematical formula for the $g$ parameter was used:

$$g = \frac{\partial E}{\partial T} = \frac{\beta_1 e^{(\beta_5+\beta_7 T+\beta_1 A+\beta_3 P+\beta_4 Income+\beta_6 Gender+\beta_6 Age)}}{[1 + e^{(\beta_5+\beta_7 T+\beta_1 A+\beta_3 P+\beta_4 Income+\beta_6 Gender+\beta_6 Age)}]^2}$$

| lnY_hat | Coef. | Std. Err. | z     | P>|z|  | [95% Conf. Interval] |
|---------|-------|-----------|-------|------|----------------------|
| T       | 11.16923 | 0.93624 | 119.30 | 0.000 | 10.98573              11.35273 |
| A       | -2.39058 | 0.1686854 | -14.17 | 0.000 | -2.721797              -2.059962 |
| P       | -0.6218471 | 0.471497 | -1.30 | 0.192 | -1.557043              0.313349 |
| income  | -3.95e-06 | 6.95e-07 | -5.68 | 0.000 | -5.31e-06              -2.59e-06 |
| gender  | 0.019361 | 0.0998134 | 1.02 | 0.307 | -0.093686              0.2975668 |
| calcage | 0.0043125 | 0.0037138 | 1.16 | 0.246 | -0.0029663             0.0115914 |
| _cons   | -3.770777 | 0.2392954 | -15.76 | 0.000 | -4.239887              -3.301966 |

sigma_u 1.264258
sigma_e 1.2007397
rho      0.52575102 (fraction of variance due to u_i)
G Media Effect Factor on the Motivation to Comply

To model the effect of media on tax morale we mainly relied on analyzing the data from our ALP survey. In our analysis of the ALP data, we looked at how the perception of evasion of the respondents amongst one hundred people like themselves changes if they knew about the evasion rates at the population level. Presented in this appendix is another analysis that we carried which looked at how people may increasingly search for information and news about tax evasion on the internet as the tax gap increases. Initially, we chose to use Twitter data to quantify this. However, this analysis was not satisfactory and instead used data from Google trends. Here we describe this approach.

We translated the term "tax evasion" into German, Italian, Greek, Danish and Swedish. These terms were used as separate search terms for the respective countries in Google Trends during the calendar year 2009 to compare the countries. This comparison generated Google Trends charts with weekly numbers. These numbers represent search interest relative to the highest point on the chart and thus do not convey absolute search volume. However, they can be used to compare interest search volume across countries. Figure G.1 shows the time series plot for the search interest relative to the highest point on the chart for the given country. Table G.1 provides the search terms for each country. The table also shows the tax gap for the different European countries for 2009 taken from the tables on pages 10-12 of the report on the European Tax Gap by Richard Murphy. Since the US tax gap for 2009 was not available we used the tax gap value from the IRS report on the Tax Gap for Tax Year 2006 Overview.

Using the data presented in Table G.1 for the mean and maximum interest, we fitted two curves. Figure G.2 shows the two curves together with the data for the mean interest for each country. We fitted a sinusoidal function taken from the functional form of CDF of a log normal distribution. Equation G.1 shows the functional form used.

\[
y = \frac{1}{2} + \frac{1}{2} \operatorname{erf} \left( \frac{s \ln x - s \ln m}{\sqrt{2}} \right), \quad (G.1)
\]

Here, \(x\) represents the tax gap expressed as a proportion and \(y\) the Google Trends mean interest value expressed as a proportion and \(\operatorname{erf}\) is the error function. The parameters \(s\) and \(m\) are found from the fitting procedure. Defining \(Y = \operatorname{inv.erf} [2y - 1]\) and \(X = \ln x\), this equation can be expressed as \(Y = kX + c\). Thus creating the transformed dataset for \(Y\) and \(X\) and by carrying out a linear regression of \(Y\) on \(X\) we can estimate the values for \(k\) and \(c\). One further assumption made in the fit is that the tax gap reaches 100%, so does the media interest. Therefore we entered a new fictitious point with \(x = 0.99\) and \(y = 0.99\) in our data set to describe this case. The parameters for \(m\) and \(s\) were extracted from the linear regression model estimates for \(k\) and \(c\) as follows: \(s = k\sqrt{2}\) and \(m = e^{-c/k}\). The best fit values are shown in figure G.2.

For low tax gap values, the mean media interest is smaller. However, in our model, we assume

---

87Google Trends: [www.google.com/trends](http://www.google.com/trends)

88Google Trends did not allow to compare all countries at one time, so we broke the search into three separate comparisons. Each of the three comparisons included Greece and two additional countries. Greece was chosen since it gave the highest search interest value and thus could be used as our base case for comparing across the three separate searches.

89We did encounter one important and strange problem with using Google Trends: The interest over time numbers and the plots change depending on the day you carry out the search. This problem occurs even for past time windows. This problem has been documented in various online blogs. For example see the blog called [Why Google Trends data change everyday?](http://www.example.com/why-google-trends-data-change-everyday). These changes, however, do not modify the ordering of the countries nor dramatically change the magnitude of average interest over the whole year 2009. Search results presented here were carried out on August 23rd, 2016.
Figure G.1: Comparison of the interest over time over the year 2009 for the search term "Tax evasion" translated into the respective language of each for different countries shown. As described in [https://support.google.com/trends](https://support.google.com/trends), “the numbers represent search interest relative to the highest point on the chart. A value of 100 is the peak popularity of the term. All the other points are relative to this point.” The highest point in these charts was for Greece during the week of January 10th, 2009. This means that the term φοροδιαφυγή + ΦΟΡΟΔΙΑΦΥΓΗ was searched the most in that particular week in Greece among the considered countries during 2009.

that the yearly media interest is stochastic and can take values from 0 to the maximum media interest found in by our Google Trends research. The value is taken using a beta-PERT distribution bounded by zero and a maximum value given by the dashed green curve and with a mean given by the dashed blue curve shown in figure G.2. We express the form of the maximum media interest by a similar functional form as that used for the mean interest. This is given in equation G.2. The difference is that we assume an extra offset term $K$ which gives the value of the maximum media interest when the tax gap is 0%.

$$y_{\text{max}} = K + (1 - K) \cdot \left\{ \frac{1}{2} + \frac{1}{2} \text{erf} \left[ \frac{s \ln x - s \ln m}{\sqrt{2}} \right] \right\}. \quad \text{(G.2)}$$

We assume the same values for $s$ and $m$ found previously, and we use the data for the peak or maximum Google Trends interest given in table G.1 to estimate $K$. Using equations G.1 and G.2 we find that $K$ can be expressed as $K = \left[ y_{\text{max}} - y \right] / (1 - y)$. So taking the fit for the mean media interest as our $y$ values and the data for the maximum media interest for our $y_{\text{max}}$ values at the tax gap points for the different countries, we estimate $K = \left[ y_{\text{max}} - y \right] / (1 - y)$. The best fit value for $K$ is shown in figure G.2. The fit can be checked by comparing the column for the max interest to the column for the max media interest in table G.1.
<table>
<thead>
<tr>
<th>Country</th>
<th>Search Terms</th>
<th>Tax Gap (%)</th>
<th>Gap mean Interest (%)</th>
<th>Gap max Interest (%)</th>
<th>Gap min Interest (%)</th>
<th>Mean Media Interest (%)</th>
<th>Max Media Interest (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Denmark</td>
<td>skatteunddragelse</td>
<td>15.00</td>
<td>13.80</td>
<td>46</td>
<td>0</td>
<td>11.00</td>
<td>43.80</td>
</tr>
<tr>
<td>Germany</td>
<td>Steuerhinterziehung</td>
<td>13.80</td>
<td>21.10</td>
<td>39</td>
<td>12</td>
<td>8.50</td>
<td>42.20</td>
</tr>
<tr>
<td>Greece</td>
<td>φοροδιαφυγή and ΦΟΡΟΔΙΑΦΥΓΗ</td>
<td>21.60</td>
<td>31.70</td>
<td>100</td>
<td>0</td>
<td>28.30</td>
<td>54.70</td>
</tr>
<tr>
<td>Italy</td>
<td>evasione fiscale</td>
<td>21.30</td>
<td>21.40</td>
<td>56</td>
<td>9</td>
<td>27.50</td>
<td>54.20</td>
</tr>
<tr>
<td>Sweden</td>
<td>skattesmitning</td>
<td>15.80</td>
<td>0.50</td>
<td>12</td>
<td>0</td>
<td>12.90</td>
<td>45.00</td>
</tr>
<tr>
<td>UK</td>
<td>tax evasion</td>
<td>11.10</td>
<td>11.10</td>
<td>20</td>
<td>3</td>
<td>3.90</td>
<td>39.30</td>
</tr>
<tr>
<td>US</td>
<td>tax evasion</td>
<td>14.50</td>
<td>12.10</td>
<td>37</td>
<td>7</td>
<td>9.90</td>
<td>43.10</td>
</tr>
</tbody>
</table>

Table G.1: Tax Gap percentage and Google Trends search terms for each country. Also shown are the Google Trends mean interest, max interest and minimum interest as a percentage of the maximum interest in the time series shown in figure G.1. The last two columns show fitted values for the mean and maximum interest that we use in the model. The form of the fit is a log-normal CDF S-Curve.

Figure G.2: Fitted curves of the Google Trends data which we use in our model to describe the media interest as a function of the tax gap. The dashed blue curve is a fit to the mean Google Trends interest values over the whole year 2009 and the dashed green line is a fit to the maximum interest values. We interpret a 100% media interest as the point where the level of media coverage on tax evasion has reached a maximum point such that the general population following the media are saturated with this information.
H Assortativity odds ratio measure

In addition to Newman’s measure of assortative mixing in a network [133], we also considered a different measure of that is based on an odds ratio that compares the frequency of edges connecting similar types of nodes in the network to the same situation if the network links were random and described by an Erdős-Renyi graph.

To explain this measure, we start by first considering a fully connected network with $N$ nodes. The number of edges is $N(N-1)/2$, or equally $\binom{N}{2}$. Each node has a color attribute and can either be red or green. Consider the case where half of the nodes are red and the other half are green. Intuitively, 25% of edges connect red to red, 50% red to green and 25% green to green. This is indeed the case because we have $\binom{N/2}{2}$ edges connecting red-red which is equal to 25% of $\binom{N}{2}$ in the limit of $N \to \infty$. Now we consider an Erdős-Renyi, not fully connected network with no bias in the way red and green nodes connect. This random network can be obtained by randomly selecting and removing a certain proportion of edges from the fully connected network. Since there is an equal probability of selecting and removing an edge we still get 25% of edges connecting red to red, 50% red to green and 25% green to green.

Generalizing, let’s say we have $N$ nodes of which $G$ are green and $R$ are red nodes ($N = G + R$) and we have $M$ edges, then if the network is fully connected $M = \binom{N}{2}$ and the proportion of red to red edges is $\binom{R}{2}/\binom{N}{2}$. As before, as we decrease the number of edges starting from $M$ in a random and unbiased way, we get an Erdős-Renyi network where the proportions of the type of edges are unaltered.

If we are given data for a network with $N$ nodes and $M$ edges which has a different structure to the Erdős-Renyi network and where there maybe assortative mixing in the way red and green nodes connect then a measure of the level of assortative mixing of how red nodes mix with other red nodes is given by

$$O_{RR/\text{RG}} = \frac{N_{\text{RR}}}{N_{\text{RG}}} \cdot \left[ \frac{\binom{R}{2}/M}{R \cdot G/M} \right]^{-1} = \frac{N_{\text{RR}}}{N_{\text{RG}}} \cdot \left[ \frac{\binom{R}{2}}{R \cdot G} \right]^{-1} = \frac{N_{\text{RR}}/N_{\text{RG}}}{E_{\text{RR/\text{RG}}}}, \tag{H.1}$$

which is a type of odds ratio between the terms $N_{\text{RR/\text{RG}}}$ and $E_{\text{RR/\text{RG}}}$. Here, $N_{\text{RR/\text{RG}}} = N_{\text{RR}}/N_{\text{RG}}$ which is the ratio of the number of edges in our network that connect red to red nodes and red to green nodes. Instead, the denominator $E_{\text{RR/\text{RG}}}$ is the ratio of edges connecting red to red nodes, with the edges connecting red to green nodes if the edges in our network were randomly redistributed. Hence, any number larger than 1 indicates assortative mixing and any number smaller than 1 indicates disassortative mixing. Thus, a value of $O_{\text{RR/\text{RG}}} = 0.5$ means that the proportion of edges connecting red to red nodes, to the edges connecting red to green nodes in our network is a half of what we would expect if the edges were instead randomly redistributed. Likewise, a value of $O_{\text{RR/\text{RG}}} = 2$ means that the proportion of edges connecting red to red nodes, to the edges connecting red to green nodes in our network is a twice as many what we would expect if the edges were instead randomly redistributed.
Testing the model dynamics for stationarity

As explained in the main text, prior to using our model to run policy experiments we have carried out model verification, validations, sensitivity analyses and calibrated the model to reproduce the observed compliance distribution, the tax gap and other known indicators about taxpayers compliance behaviors. In order to carry out these verifications and analyses and to compare model case runs against each other, we have had to determine whether each case run had reached a state of stationarity and dynamics equilibrium.

During the dynamics, the model tracks the overall gross tax gap $\Xi_{t}^{\{\text{gross}\}}$, and in particular by how much it change from one year to the successive year, which here we denote as $\Delta \Xi_{t}^{\{\text{gross}\}} = \Xi_{t}^{\{\text{gross}\}} - \Xi_{t-1}^{\{\text{gross}\}}$. Specifically, the model keeps a record of the values of the gross tax gaps produced in the last $t_{\text{window}}$ years and uses them to compute the values of $\Delta \Xi_{t}^{\{\text{gross}\}}$. It then finds $\langle \Delta \Xi_{t}^{\{\text{gross}\}} \rangle_{t_{\text{window}}}$ which denotes the moving average of the first order changes in the gross tax gap over the last $t_{\text{window}}$ years. In a similar manner, the model also tracks the second order changes in the overall gross tax gap, which here is denoted as $\Delta^{2} \Xi_{t}^{\{\text{gross}\}} = \Delta \Xi_{t}^{\{\text{gross}\}} - \Delta \Xi_{t-1}^{\{\text{gross}\}}$, and finds it’s moving average $\langle \Delta^{2} \Xi_{t}^{\{\text{gross}\}} \rangle_{t_{\text{window}}}$. To test for equilibrium, we compare the values of the moving averages of the first and second order changes in gross tax gap to two different maximum tolerance values. If both moving averages are within their respective tolerance range, the model considers the dynamics to have reached a stage of dynamic equilibrium.

For our case runs, we have set the tolerance ranges to be very stringent. Specifically, we use consider $t_{\text{window}} = 200$ years and the tolerance values to be $5 \times 10^{-5}$ and $5 \times 10^{-6}$ respectively. Hence, each case run will run for at least 200 years before the model begins to test for equilibrium. However, when the model starts from very artificial initial conditions, such as setting every taxpayer to be initially compliant, and when it uses our population network with just over 1000 taxpayers (i.e., our PN1 network), the model may run for as long as 300 or 400 years and sometimes for even longer before our stringent specification for equilibrium are satisfied. The reason the model runs for so long before reaching equilibrium is that when we consider the model with just 1000 taxpayers, stochastic variability in the dynamics is high and this could lead to large deviations from the current moving average values. Consequently, this postpones the time when these conditions are satisfied. When we consider our larger population and network PN10, the time to reach equilibrium is reduced considerably.

\[\text{From a statistical physics perspective, the system is not externally driven and thus, in this case, stationarity implies a dynamic equilibrium. The equilibrium is dynamic because although at the aggregated level quantities are stationary and only randomly fluctuated about their mean value, at the individual-level each taxpayer is instead of changing behavior.}\]
J  CART and Random Forest methods

To carry out our uncertainty and sensitivity analysis, we used two supervised machine learning
algorithm that are commonly applied to analyze Big Data, namely Classification And Regression
Trees (CART) and Random Forest methods \cite{28,29}. For each model output of interest, we used
CART to create a binary decision tree. This process recursively splits the space of input parameter
values into regions that produce comparable output. As a first step, the algorithm uses regressions
to find which input parameter and its range in values best explains the variance of the model out-
put across all the model case runs. Then, using a cost function it finds the threshold value of the
input parameter that best splits the output data into two distinct groups. Using the CART ter-
minology, this creates the root-node and two first-generation leaf-nodes \footnote{If the input parameter that is considered for the splitting is numerical and continuous then the algorithm uses this regression-based method. If instead the input parameter is categorical then a different approach is used where a Gini index function is used instead of a cost function. This index provides an indication of how “pure” a leaf-node is in terms of what different categorical values of the input it includes.}. During this process
all input parameters and all possible split point values are evaluated \footnote{For a given input parameter and split point, the algorithm first splits the data into two separate training samples and runs a regression model for each sub-sample. For each of the two sub-samples, it then uses the sum of the squared difference between the training sub-sample data and the prediction from the regression to generate the cost function. The input parameter and the split point that minimizes the cost function is then chosen. Thus, the cost function is minimized when the input space is split into two regions that produce distinct outputs.}. This is then repeated for
each of these two groups or leaf-nodes. Thus, using the subset of model case runs belonging to
each group, we repeat the process and find which input parameters can further best split the out-
put data into additional second-generation leaf-nodes. This processes continues until a stopping
criteria is satisfied. The most common stopping procedure is to specify a minimum number of
model case runs that need to be assigned to each leaf node. When the method splits the model
cases runs belonging to a given node into two leaf-nodes and either leaf-node has less cases than
some minimum threshold number, then the split is not accepted and the node is taken as a final
leaf-node. Alternatively, a stopping criteria can be specified based on when the variance of the
output across all model cases contained a given leaf-node goes below a certain threshold value.

CART decision trees produce distinct regions in the input space that are separated by very
clear-cut boundaries. However, if the tree is grown very deeply, there is a risk of splitting the input
space into too many regions, some of which contain little data. Moreover, the position of these
boundaries may well depend on outliers and other anomalies in the data. By over-interpreting
the accuracy of these clear-cut boundaries we risk categorizing model case runs incorrectly. For
example, a model case run may incorrectly be categorized as belonging to one region when in
reality it is in the general vicinity of the region and very close to the region boundary. Thus,
we could risk drawing conclusions too finely from the data that we have and hence overfitting
the statistical model. Random forest (RF) algorithms address this limitation. A RF generates
a multitude of randomly selected sub-samples of the full data set of model case runs and uses
them to create an ensemble (i.e., forest) of different CART trees. Each of the decision trees will
have slightly different boundaries between the different regions. By averaging over these trees
we generate regions in input space that is no longer split by clear-cut boundaries. Moreover, the
position of these boundaries is less sensitive to the outliers and other anomalies in the data. This
is because outliers will be present in a just a few decision trees and are likely to be split amongst
different decision trees. Instead, data that is more general will be present in the vast majority of
the decision trees. In this way, the boundaries separating the different regions are smoothed over
by the averaging process. Therefore, RF provides a method for drawing conclusions that are more
robust than a single CART decision tree.
K Code and Data Repository

During the development of our model, we used RAND’s internal Gitlab server at code.rand.org for version control. RAND’s Gitlab is not externally accessible. However, in compliance with deliverables for NSF, we have made our code available externally on a public git server and this can be accessed at https://github.com/R-Vardavas/RIBSS_tax_evasion_ABM. The repository does not include the primary and secondary raw data sets used to inform the model. It, however, includes the settings and attributes of our synthetic population that was created using these datasets. All secondary datasets we used are publicly available or can be obtained by contacting the authors of the various research we used. Our primary dataset includes the survey data obtained from RAND’s American Life Panel. The survey data is freely available by logging into the ALP site at https://alpdata.rand.org/index.php?page=data&p=showsurvey&syid=456.

K.1 Repository’s Directory Structure

- Prototype.R This is used for running the model just to test out basic changes to the code. This is not the most efficient way to run the code, but if someone is interested in analyzing how the model works, going through this file line by line would be the most useful way to do it.

- TaxModel.R This file contains the main function tax.model(). The simulation is set up to run if the tax.model() is called from the R-console, after sourcing the Tax.Model.R. The default configuration that this run will use is stored as a model.config.csv file in the Inputs folder. Giving different paths will run different scenarios. One can also directly supply the config data frame as input instead of providing the path name to the file containing the configuration settings. All the inputs needed for the model, are in general, available in the Inputs folder. All the helper functions needed for the tax model to run are saved in the Library folder.

- Sensitivity Analysis All the code that was used to carry out the sensitivity analysis and calibration of the model is placed in this folder. calibration.analysis.R and sensitivity.analysis.R are the files that assessed the data after running multiple cases of the tax model with varying configurations. To run the model with varying configurations, we used calibration.analysis.wrapper.R, and sensitivity.analysis.wrapper.R. These are set up in such a way that the tax model can be run for multiple cases that have been sampled using the Latin Hypercube Sampling method.

- PolicyExperiments This has code for running all the various policy experiments we tried out using the tax model. Each experiment has been coded in a separate file. We use the runner_exp.R to run the model for the various experimental scenarios and save the data in appropriate folders. Most experiments have the model.type option to specify 1k population or 10k population agents to run the model on.

K.2 Quick Start to Running the Model

In order to run the model, the user first needs to install R and RStudio. Both these software are open source and can be downloaded from the following locations:

- R (version 3.3 or higher): https://www.r-project.org/
Once this software is installed, the steps to follow are:

1. Run RStudio and create a new project using the File menu option.

2. Clone the repository select Version Control → Git and paste the repository URL: https://github.com/R-Vardavas/RIBSS_tax_evasion_ABM.git. Figure K.1 shows the type option you should select when creating your project. This should take some time in cloning the repository to your machine.

3. Install the all the libraries needed to run the tool, please open the file prerun.R and run it using the Source option on the top right side of the RStudio tool. This should install all the libraries required for the tool.

4. To run the tool open one of the following files: ui.R or server.R, or global.R and hit the RunApp button on the top right. Ideally, at this stage, the tool will be running.
L Interactive Tool

We developed an interactive tool for our model using R Shiny. The purpose of our interactive tool was to assist us with model verification and to help new interested researchers familiarize with the features of the model. The tool preloads one of our calibrated model runs that has reached stationarity and the associated input parameters. The user can then choose to run the model forward at the status-quo settings or choose to change policy and behavioral parameters. Figure L.1 shows the policy levers tab of the tool. From here and with the slider bars, users can change the general policy levers specified by the overall audit rate, the detection efficiency, the penalty rate and the change in overall effective tax rate in percentage points (pp). Users can also specify more detailed fiscal and deterrence policies. The central panel allows users to specify the tax brackets and marginal tax rates that apply to taxpayers belonging to different filing statuses. The column named +Delta shows the actual marginal tax rate once we factor in (i) any changes in the user’s specification of the tax brackets and marginal tax rates in the central panel, and (ii) the change percentage points of the overall effective tax rate specified by the slider bar in the left-side panel. The tool then accounts for these changes and displaces an effective weighted average tax rate that applies at the population-level. Similarly, the right-side panel specifies the deterrence strategy. Users can change the effective audit rate that applies to the different income categories. As before, the column named +Delta displaces the actual audit rate once the change in audit rate specified by the slider-bar in the left-side panel is accounted for.

Figure L.2 shows the three panels with slider bars where the user can change the value of various important model parameters that control taxpayer’s compliance behaviors. Parameters have been categorized into three main groups shown by the three panels.

Figure L.1: Policy inputs tab. Here users can make general changes to the fiscal and deterrence policies in the left-side panel using the slider bars, or provide more detailed specifications using the central and right-side panels.
Figure L.2: Behavioral parameters tab. Here users can change the values of certain input parameters considered by the model. The value of each input model parameter can be changed using the slider bar and varies within a given range that was considered by our uncertainty analysis in developing the model.

Once the inputs and the policy levers have been changed or left as they are, the tool can be used to run an additional 100 year of dynamics starting from our status-quo conditions. Results of the model runs can be visualized using a set of different panel windows. A panel window can be selected using the black panel menu area on the left side. Selecting the Model Plots option will change the view and users will be presented with a similar window as that illustrated in Figure L.3. This window has two panels, each with a pull-down menu from which users can select different plots of the model run. This window can be used to compare plots against each other to gain a better sense of the model results. These plots mainly focus on individual-level tax compliance behavior and include histograms and trajectories of compliance behavior. Selecting the Marco trends option will change the view and users will be presented with a different window that shows and allows users to compare different macro-level effects. Figure L.4 illustrates the types of plots that this window displays.

Figures L.5 and L.6 show illustrate the plots produced when selecting the Micro Trends option. In the left-side panel of Figure L.5 users can visualize the entire population and social network of contracts during the different years in the model dynamics. By clicking on a taxpayer in the network, users can visualize on the right-side panel the compliance trajectory of the taxpayers, and of his/her alters on the network. The tool can also plot the average compliance behavior of all the alters belonging to the selected taxpayer. This allows users to understand at the dynamical and micro-level the effects produced by the model of social network tax compliance influences on the compliance trajectory of a given taxpayer. By double-clicking on a given taxpayer, users are presented with an egocentric network plot that focuses host on the taxpayer and his immediate social network. Figure L.6 illustrates the egocentric network plot. These egocentric network visu-
Figure L.3: The Model Plots window can be used to compare different types of plots against each other.

Figure L.4: The Macro Plots window can be used to compare different types of plots of population-level and macro economic outcomes.
Figure L.5: The Micro Plots allows users to visualize the entire network of taxpayers. The slider bar allows users to see the dynamics on the network. The network visualization emphasizes the taxpayers who get audited but not penalized (green stars) and taxpayers who get audited and penalized for tax evasion (large white circles).

Alizations can be more helpful than visualizing the entire network and be used best in congestion with the compliance trajectory plots shown in the high-side panel.
Figure L.6: The Micro Plots also allows users to visualize the egocentric networks of taxpayers showing their compliance behavior and that of their alters.
References


Ref–6


Ref–9


Ref-11