



Tax Administration
Research Centre

Tax evasion on a social network

TARC DISCUSSION PAPER 022 - 18

DUCCIO GAMANNOSSI DEGL'INNOCENTI, TARC, UNIVERSITY OF EXETER
MATTHEW D RABLEN, UNIVERSITY OF SHEFFIELD

Tax Evasion on a Social Network*

Duccio Gamannossi degl’Innocenti[†]
d.gamannossi@exeter.ac.uk

Matthew D. Rablen[‡]
m.rablen@sheffield.ac.uk

October 12, 2018

Abstract

We relate tax evasion behavior to a substantial literature on social comparison in judgements. Taxpayers engage in tax evasion as a means to boost their expected consumption relative to others in their social network. The unique Nash equilibrium of the model relates optimal evasion to a (Bonacich) measure of network centrality: more central taxpayers evade more. Given that tax authorities are now investing heavily in big-data tools that aim to construct social networks, we investigate the value of acquiring network information. We do this using networks corresponding closely in structure to those observed empirically. In particular, we allow for celebrity taxpayers, whose consumption is widely seen, and who are systematically of higher wealth. We show that there are pronounced returns to the initial acquisition of network information, albeit targeting audits with highly incomplete knowledge of social networks may be counterproductive.

JEL Classification: H26, D85, K42.

Keywords: Tax evasion, Social networks, Network centrality, Optimal auditing, Social comparison, Relative consumption.

*Acknowledgements: We thank Kangoh Lee, Maria Gil Molto, Jo McHardy, and participants at the 6th TARC Annual Conference (Exeter), PET18 (Hue), Tax Research Network (Birmingham) and SIEP (Padua) for helpful comments on versions of the manuscript. We acknowledge the ESRC for financial support (ES/S00713X/1).

[†]Tax Administration Research Centre (TARC), University of Exeter Business School, University of Exeter, Exeter, EX4 4PU, UK.

[‡]Department of Economics, University of Sheffield, Sheffield, S1 4DT, UK.

1 Introduction

Tax evasion is a significant economic phenomenon. Estimates provided by the UK tax authority put the value of the tax gap – the difference between the theoretical tax liability and the amount of tax paid – at 6.5 percent (H.M. Revenue and Customs, 2016). Academic studies for the US and Europe put the gap substantially higher, at around 18-20 percent (Cebula and Feige, 2012; Buehn and Schneider, 2016).

In this paper we link evasion behavior to a mass of evidence that people engage continually in comparisons with others (social comparison). Utility, evidence for developed economies suggests, is in large part derived from consumption relative to social comparators, rather than from its absolute level (e.g., Ferrer-i-Carbonell, 2005; Luttmer, 2005; Clark and Senik, 2010; Mujcic and Frijters, 2013). The evolutionary processes that might explain this phenomenon are explored in Postlewaite (1998), Rayo and Becker (2007) and Samuelson (2004), among others. Researchers have proposed that social comparison can explain economic phenomena including the Easterlin paradox (Clark *et al.*, 2008; Rablen, 2008), stable labor supply in the face of rising incomes (Neumark and Postlewaite, 1998); the feeling of poverty (Sen, 1983); the demand for risky activities (Becker *et al.*, 2005); and migration choices (Stark and Taylor, 1991). There are important consequences for consumption and saving behavior (Harbaugh, 1996; Hopkins and Kornienko, 2004) the desirability of economic growth (Layard, 1980, 2005), and for tax policy (Boskin and Sheshinski, 1978; Ljungqvist and Uhlig, 2000).

We provide a network model in which taxpayers are assumed to have an intrinsic concern for consumption relative to that of other “local” taxpayers with whom they are linked on a social network.¹ In this context, taxpayers may seek to evade tax so as to improve their standing relative to those they compare against. The model exhibits strategic complementarities in evasion choices, so that more evasion by one taxpayer reinforces other taxpayers’ decisions to evade also. Following the lead of Ballester *et al.* (2006), we utilize linear-quadratic utility functions to provide a characterization of Nash equilibrium. We show that there is a unique Nash equilibrium in which evasion is a weighted network centrality measure of the form proposed by Bonacich (1987). Network centrality is a concept developed in sociology to quantify the influence or power of actors in a network. It counts the number of all paths (not just shortest paths) that emanate from a given node, weighted by a decay factor that

¹The economics of networks is a growing field. For recent overviews, see Ioannides (2012), Jackson and Zenou (2015), and Jackson *et al.* (2017). Our analysis connects to a broader literature that applies network theory to the analysis of crime (e.g., Glaeser *et al.*, 1996; Ballester *et al.*, 2006).

decreases with the length of these paths. In this sense, our contribution combines sociological and economic insights in seeking to understand tax evasion behavior.

Although the model is simple enough to admit an analytic solution, it is also sufficiently rich that it may be used to address a range of questions of interest to academics and practitioners in tax authorities. Here we focus on two such questions: first, we investigate – for an arbitrary network structure – how changes in the model’s exogenous parameters affect optimal evasion. Second, in the light of growing investment by tax authorities into “big data” tools that seek to construct the underlying network, we investigate the value to a tax authority – in terms of additional revenue raised through audits – of knowing the structure of the social network. The analysis is performed on a class of generative networks that possess many of the empirically observed features of social networks – in particular allowing for highly visible celebrity taxpayers. Our results compare audit revenue outcomes when audits are targeted using the available network information with audit revenues under random auditing. We show that there are strong returns to a tax authority from moving from not observing the social network at all to observing around 20 percent of the links. Attempting to target audits with a very limited understanding of the social network is, however, shown to actually be counterproductive. A greater concentration of links within a social network increases the value of possessing at least some network information. These findings are robust to the presence or absence of unobserved preference heterogeneity.

An important feature of our model is that it addresses explicitly the role of *local* comparisons on a social network. By contrast, the existing analytical literature on tax evasion allows only *global* (aggregate) social information to enter preferences: the global statistic that taxpayers are assumed to both have a concern for, and to be able to observe, is either (i) the proportion of taxpayers who report honestly (Gordon, 1989; Myles and Naylor, 1996; Davis *et al.*, 2003; Kim, 2003; Traxler, 2010; Ratto *et al.*, 2013); (ii) the average post-tax consumption level (Goerke, 2013); (iii) the level of evasion as a share of GDP (Dell’Anno, 2009); or (iv) the average tax payment (Mittone and Patelli, 2000; Panadés, 2004).

While reducing social information to a single statistic known to all taxpayers promotes analytical tractability, it is problematic in other respects. First, from the perspective of modelling with explicit social networks, assuming that taxpayer’s observe aggregate-level information is implicitly the assumption that every taxpayer observes every other taxpayer. When, as we suppose, observability is only between linked taxpayers, full observability is

equivalent to the assumption that the social network is the complete network (in which every taxpayer is directly linked to all other taxpayers). But there are reasons to think that relative consumption externalities are, in fact, heterogeneous across individuals. In particular, we know that people’s reference group is typically composed of “local” comparators such as neighbors, colleagues, and friends (Luttmer, 2005; Clark and Senik, 2010).² Moreover, implicitly assuming a complete network implies that all taxpayers are equally connected socially, thereby ruling out, in particular, the existence of “stars” or “celebrities” whose consumption is very widely observed in the network. Yet, this feature of social networks may matter for the targeting of tax audits (Andrei *et al.*, 2014).

The only literature that has enriched the analysis of social information to allow for local comparisons is that which uses agent-based simulation techniques as an alternative to analytical methods. Models in this tradition nonetheless employ representations of social networks that appear to differ markedly from real world examples. A common property of the network structures employed (e.g., Korobow *et al.*, 2007; Hokamp and Pickhardt, 2010; Bloomquist, 2011; Hokamp, 2014) is that the number of taxpayers who observe each taxpayer is fixed, thereby ruling out the existence of highly-observed celebrity taxpayers. Other authors (e.g., Davis *et al.*, 2003; Hashimzade *et al.*, 2014, 2016) utilize an *undirected* network, meaning that, if i is linked to j , then necessarily j is linked to i . Yet social networks display strong asymmetry in the direction of links (Foster *et al.*, 2010; Szell and Thurner, 2010).³ We offer a model that is both analytically tractable and that allows for local comparisons on an arbitrary social network. In this sense, our approach lies in the cleavage between existing analytical and agent-based approaches, and is complementary to each.⁴ We perform simulation analysis on a class of generative networks that are not subject to the restrictions discussed above, and which are widely utilized to model network structures in the natural sciences. Our methodology in this regard, therefore, has applicability beyond the current context of tax evasion.

In related research, Goerke (2013) assumes an intrinsic concern for relative consumption

²More generally, relative consumption externalities may be viewed as a form of *peer effect*. In other contexts, generative models of peer effects predict heterogeneous exposure. For instance, when job information flows through friendship links, employment outcomes vary across otherwise identical agents with their location in the network of such links (Calvó-Armengol and Jackson, 2004).

³Zaklan *et al.* (2008) and Andrei *et al.* (2014) are among exceptions that do explore more general network structures.

⁴By extending analytical understanding of network effects upon tax evasion – in particular being able to prove formal comparative statics properties of the model – we assist the interpretation of simulation output from related agent-based models.

by taxpayers. The primary focus of his contribution is, however, the derived impact on tax evasion from endogenous changes in labor supply, whereas we treat earned income as an exogenous parameter. In the remaining literature that considers a social dimension to the tax evasion decision, taxpayers are assumed to derive utility solely from absolute consumption, but react nonetheless to social information because they experience social stigma – the extent of which depends on the evasion of other taxpayers – if caught evading. The focus of much of this literature is on the potential for multiple equilibria, whereas our model yields a unique equilibrium. While a concern for relative consumption is compatible with the simultaneous existence of social stigma towards evaders, the two approaches differ in emphasis. Underlying the idea of social stigma is the concept of *social conformity*, in which individuals seek to belong to the crowd, whereas the presumption of relative consumption theories is that individuals seek to stand out from the crowd. A literature relating to this point in the context of tax evasion offers strong evidence that social information impacts compliance behavior (Webley *et al.*, 1988; De Juan *et al.*, 1994; Alm and Yunus, 2009; Alm *et al.*, 2017), but rejects social conformity as the underlying mechanism (Fortin *et al.*, 2007).

The paper proceeds as follows: section 2 develops a formal model of tax evasion on a social network. Section 3 analyzes the comparative statics of optimal evasion with respect to information transmitted through the social network. Section 4 considers the value of network information to a tax authority, and section 5 concludes. Proofs are collected in the Appendix, and figures are at the very rear.

2 Model

Let \mathcal{N} be a set of taxpayers of size N . A taxpayer $i \in \mathcal{N}$ has an (exogenously earned) income $W_i > 0$. By law taxpayers should declare W_i to the tax authority and pay tax $\theta(W_i)$, where $\theta : \mathbb{R}_{\geq 0} \mapsto (0, W_i)$ is a non-decreasing function. If a taxpayer declares their true gross income, W_i , they receive a (legal) net disposable income $X_i \equiv X_i(W_i) = W_i - \theta(W_i)$. Taxpayers can, however, choose to declare less than their true income, thereby evading an amount of tax $E_i \in [0, W_i - X_i]$. Taxpayer i is audited with probability $p_i \in (0, 1)$ in each period. Heterogeneity in the p_i can arise, for example, if the tax authority conditions audit selection upon observable features of taxpayers. Audited taxpayers face a fine at rate $f > 1$ on all undeclared tax, à la Yitzhaki (1974).

Taxpayers are assumed to derive utility from their level of consumption relative to a reference

level of consumption R_i (the determination of which we shall discuss later).

In each period, taxpayers behave as if they maximize expected utility, where the utility of taxpayer i is denoted by $U_i(\cdot)$. Most of our analysis shall be undertaken under the assumption that $U_i(\cdot)$ is of the linear-quadratic form, but the model is more parsimoniously outlined for the more general case. The expected utility of taxpayer i is therefore given by

$$\mathbb{E}(U_i) \equiv [1 - p_i] U_i(C_i^m - R_i) + p_i U_i(C_i^a - R_i), \quad (1)$$

where consumption in the audited state (C_i^a) and not-audited state (C_i^m) is given by:

$$C_i^m \equiv X_i + E_i; \quad C_i^a \equiv C_i^m - fE_i. \quad (2)$$

An obvious objection to this formulation is that it neglects entirely the possibility of absolute utility. Although an absolute component to utility surely exists, we note that measures of subjective wellbeing typically become uncorrelated with absolute income above a threshold of average national income estimated at \$5,000 (in 1995, PPP) by Frey and Stutzer (2002). As most citizens of developed countries lie above this threshold, our model may be a reasonable approximation in such cases. Optimal evasion is the solution to the problem $\max_{E_i} \mathbb{E}(U_i)$ subject to the Cournot constraint that reference consumption, R_i , is taken as given. The first order condition for optimal evasion is therefore given by

$$[1 - p_i] U_i'(C_i^m - R_i) - p_i f U_i'(C_i^a - R_i) = 0. \quad (3)$$

2.1 Reference Consumption

Reference consumption, R_i , is a function of social comparison. To formalize the notion of social comparison, we assume that each taxpayer observes the consumption of a non-empty set of taxpayers $\mathcal{R}_i \subset \mathcal{N}$, a set we term the *reference group*. A taxpayer, i , is said to be *observed* if their consumption is visible to at least one other taxpayer in the network, i.e., $i \in \cup_{j \in \mathcal{N} \setminus i} \mathcal{R}_j$.

We represent the observability of consumption in the form of a directed network (graph), where a link (edge) from taxpayer (node) i to taxpayer j indicates that i observes j 's consumption. Links are permitted to be subjectively weighted, for some members of the reference group may be more focal comparators than are others. The network is represented as an $N \times N$ (adjacency) matrix, \mathbf{G} , of subjective comparison intensity weights $g_{ij} \in [0, 1]$, where

$g_{ii} = 0$. We normalize the g_{ij} for each taxpayer to sum to unity: $\sum_{j \in \mathcal{R}_i} g_{ij} = 1$. Taxpayers i and j with $g_{ij} > 0$ are said to be *linked*. Accordingly, the reference group of taxpayer i is the set of all taxpayers to whom i is linked: $\mathcal{R}_i = \{j \in \mathcal{N} : g_{ij} > 0\}$. For later reference, a network, \mathbf{G} , in which there is a path (though not necessarily a direct link) between every pair of taxpayers is said to be *connected*.

Writing expected consumption as $q_i = [1 - p]C_i^n + pC_i^a = X_i + [1 - p_i f]E_i$, we define \bar{q}_i as the (weighted) mean of q_j over i 's reference group ($j \in \mathcal{R}_i$). This weighted mean is written conveniently as $\bar{q}_i \equiv \bar{q}_i(\mathbf{q}) = \mathbf{g}_i \mathbf{q}$, where \mathbf{g}_i is the i^{th} row of \mathbf{G} , and \mathbf{q} is a $N \times 1$ vector of the expected consumptions. We then set

$$R_i = \bar{q}_i. \quad (4)$$

2.2 Nash Equilibrium

Using (4) in the first order condition (3), we now solve for the unique Nash equilibrium of the model. To do this, we first define a notion of network centrality due to Bonacich (1987), which computes the (weighted) discounted sum of paths originating from a taxpayer in the network:

Definition 1 *Consider a network with (weighted) adjacency matrix \mathbf{A} . For a scalar β and weight vector $\boldsymbol{\alpha}$, the weighted Bonacich centrality vector is given by $\mathbf{b}(\mathbf{A}, \beta, \boldsymbol{\alpha}) = [\mathbf{I} - \beta \mathbf{A}]^{-1} \boldsymbol{\alpha}$ provided that $[\mathbf{I} - \beta \mathbf{A}]^{-1}$ is well-defined and non-negative.*

In Definition 1, the scalar β specifies discount factor that scales down (geometrically) the relative weight of longer paths, while the vector $\boldsymbol{\alpha}$ is a set of weights. In the present context the matrix $[\mathbf{I} - \beta \mathbf{A}]^{-1}$ is a form of social comparison multiplier. It measures the way in which actions by one taxpayer feed through into other taxpayers' actions. Ballester *et al.* (2006) show that $[\mathbf{I} - \beta \mathbf{A}]^{-1}$ will be well-defined, as required in Definition 1, when $1 > \beta \rho(\mathbf{A})$, where $\rho(\mathbf{A})$ is the spectral radius of \mathbf{A} . In our context, this condition is that the local externality that a taxpayer's evasion imparts upon other taxpayers cannot be too strong. If local externality effects are too strong then the set of equations that define an interior Nash equilibrium of the model have no solution. In this case, multiple corner equilibria can instead arise (see, e.g., Bramoullé and Kranton, 2007). Focusing on the case when local externality effects are not too strong, we have the following Proposition:

Proposition 1 *If*

(i) utility is linear-quadratic, $U_i(z) = [b_i - \frac{a_i z}{2}] z$, with $a_i \in (0, \frac{b_i}{W_i})$ and $b_i > 0$ for all $i \in \mathcal{N}$;

(ii) $1 > \rho(\mathbf{M})$; $[\mathbf{I} - \mathbf{M}] \theta(\mathbf{W}) - \boldsymbol{\alpha} > \mathbf{0}$;

then there is a unique interior Nash equilibrium, at which the optimal amount of tax evaded is given by

$$\mathbf{E} = \mathbf{b}(\mathbf{M}, \mathbf{1}, \boldsymbol{\alpha}),$$

where

$$\begin{aligned} m_{ij} &= \frac{[1 - p_i f][1 - p_j f]}{\zeta_i} g_{ij}; & \zeta_i &= [1 - p_i f]^2 + p_i [1 - p_i] f^2 > 0. \\ \alpha_{i1} &= \frac{1 - p_i f}{a_i \zeta_i} \{b - a_i [X_i - R_i(\mathbf{X})]\}; \end{aligned}$$

Proposition 1 characterizes the unique interior Nash equilibrium of the model, and the conditions under which it arises, for the case of linear-quadratic utility. The restrictions in (i) on the parameters $\{a_i, b_i\}$ are sufficient to ensure that marginal utility is everywhere positive. The restrictions in (ii) guarantee an interior equilibrium: the first ensures $E_i > 0$, while the second ensures $E_i < \theta(W_i)$, so that the amount a taxpayer evades does not exceed the amount of tax they owe. The uniqueness of equilibrium evasion follows intuitively from the observation that, under linear-quadratic utility, each taxpayer's best response function is linear in the evasion of every other taxpayer.

According to Proposition 1 a taxpayer's optimal evasion corresponds to a Bonacich centrality on the social network \mathbf{M} , weighted to reflect a taxpayer's marginal utility of consumption.⁵ By this measure, *taxpayers that are more central in the social network evade more*. The social network \mathbf{M} transforms the underlying comparison intensity weights, g_{ij} , by a factor $[1 - p_i f][1 - p_j f]\zeta_i^{-1} \in (0, 1)$ that reflects potential heterogeneity in audit probabilities across taxpayers. It follows that, in the special case that all taxpayers face a common audit probability, as might occur if a tax authority has committed to a policy of random auditing, no adjustment to the underlying comparison intensity weights is warranted. In this case, therefore, optimal evasion is a weighted Bonacich centrality measure on the untransformed network \mathbf{G} :

⁵Marginal utility in the linear-quadratic specification is given by $U'(z) = b - az$. Accordingly, the term in braces in the expression for α_{i1} is the marginal utility from one's own legal consumption, X_i , relative to a reference level of consumption reflecting the weighted average over the reference group of legal consumption.

Corollary 1 *Under the conditions of Proposition 1 and setting $p_i = p$ for all $i \in \mathcal{N}$, the unique interior Nash equilibrium for evasion is given by $\mathbf{E} = \mathbf{b}(\mathbf{G}, \omega, \boldsymbol{\alpha})$, where*

$$\omega = \frac{[1 - pf]^2}{\zeta} < 1.$$

What if utility is not linear-quadratic? For an arbitrary twice-differentiable utility function we may generalize the model by considering the first order linear approximation around a Nash equilibrium to a set of (potentially non-linear) first order conditions of the form in (3). The resulting set of equations are given by

$$\mathbf{E} = \mathbf{J}\mathbf{E} + \hat{\boldsymbol{\alpha}} = [\mathbf{I} - \mathbf{J}]^{-1} \hat{\boldsymbol{\alpha}} = \left[\sum_{k=0}^{\infty} \mathbf{J}^k \right] \hat{\boldsymbol{\alpha}}, \quad (5)$$

where $\hat{\boldsymbol{\alpha}}$ is again a vector of weights for the different taxpayers, and \mathbf{J} is a matrix of coefficients measuring how actions interact. By appropriate decomposition of \mathbf{J} , therefore, a solution to the equation system in (5) is a Bonacich centrality measure of the form contained in Definition 1.

The model we have outlined is sufficiently rich that it may be used to address a range of questions of interest to academics and practitioners in tax authorities. Subsequent sections will consider how social information carried by the network affects optimal evasion, and the value of knowing the network to a tax authority.

3 Comparative Statics of Network Interaction

We now use the model in Proposition 1 to understand the way in which social information affects the evasion decision. We do this for an arbitrary social network, subject to condition (ii) in Proposition 1 holding.

A basic property of the model is that, under the assumptions of Proposition 1, the model exhibits strategic complementarities in evasion choices: an increase in evasion by one taxpayer induces others to do likewise.⁶ This is equivalent to the expected utility of taxpayer i being

⁶The version of the model we present here is simple enough to admit computation of exact comparative statics. An advantage of the strategic complementary property, however, is that, if one only seeks the signs of the comparative statics (as might be appropriate in more complex versions of the model where the computation of exact comparative statics is burdensome), these can be elucidated straightforwardly using the theory of monotone comparative statics (e.g., Edlin and Shannon, 1998; Tremblay and Tremblay, 2010).

supermodular in the cross evasion choices of another taxpayer j belonging to i 's reference group:

$$\frac{\partial^2 \mathbb{E}(U_i)}{\partial E_i \partial E_j} = a_i g_{ij} [1 - p_i f] [1 - p_j f] > 0 \quad j \in \mathcal{R}_i.$$

We now analyze how the evasion decision of a taxpayer i , E_i , is affected by a permanent marginal increase in a parameter z_j belonging to a different agent $j \neq i$. Differentiating the expression for evasion in Proposition 1 we obtain:

Proposition 2 *Under the conditions of Proposition 1 it holds at an interior Nash equilibrium that:*

$$\begin{aligned} \frac{\partial E_i}{\partial W_j} &= b_{1i} \left(\mathbf{M}, 1, \frac{\partial \boldsymbol{\alpha}}{\partial X_j} \right) \geq 0; \\ \frac{\partial E_i}{\partial p_j} &= b_{1i} \left(\mathbf{M}, 1, \frac{\partial \mathbf{M}}{\partial p_j} \mathbf{E} + \frac{\partial \boldsymbol{\alpha}}{\partial p_j} \right) \leq 0. \end{aligned}$$

The results in Proposition 2 underscore that the attributes of other taxpayers, and the treatment of other taxpayers by the tax authority, both affect own compliance. Moreover, the precise effects are heterogeneous across taxpayers, depending upon how “close” taxpayers are in the social network. In respect of sign, these results are in line with those of models of tax evasion that introduce social concerns through a social norm for compliance, albeit there are important differences in economic interpretation.

The first result is that an increase in the income of taxpayer j induces taxpayer i to evade more. When j gets richer this pushes up their consumption, causing those taxpayers who observe j 's consumption to feel poorer in relative terms. This, in turn, induces these taxpayers to increase their evasion in an attempt to boost their consumption. This behavior, in turn, induces yet a further set of taxpayers to also feel poorer, and also increase their evasion, and so on. If the network M is connected then this ripple effect reaches every taxpayer in the network, so the result in Proposition 2 may be strengthened to a strict inequality. Nonetheless, the effect is on average greater for those taxpayers who are closest to j in the network, especially those with a direct link to j . Extending this argument, if the network M is not connected, then there exists at least one taxpayer pair $\{i, j\}$ between whom social information does not flow. For such $\{i, j\}$ pairs it will hold that $\partial E_i / \partial W_j = 0$.

The second result in Proposition 2 is that the evasion of taxpayer i responds negatively to the level of tax authority enforcement of other taxpayers in the social network. When a

taxpayer j experiences an increase in audit probability they decrease their evasion. This decreases the evasion required of taxpayer i to maintain a given level of expected relative consumption, leading i to evade less. The result can be strengthened to a strict inequality if the network M is connected.

4 Audit Targeting and Network Structure

Can tax authorities observe links in social networks? Although surely the full gamut of links cannot be observed, importantly, there exist some individuals – celebrities – for whom it is common knowledge that many people observe them. Also, even for non-celebrities, the idea that tax authorities know at least something about people’s associations is becoming more credible with the advent of “big data”. The UK tax authority, for instance, uses a system known as “Connect”, operational details of which are in the public domain (see, e.g., Baldwin and McKenna, 2014; Rigney, 2016; Suter, 2017). Connect cross-checks public sector and third-party information, seeking to detect relationships among actors. According to Baldwin and McKenna (2014), the system produces “spider diagrams” linking individuals to other individuals and to legal entities such as “property addresses, companies, partnerships and trusts.” The IRS is known to have also invested in big data heavily, but has so far been much more reticent in revealing its capabilities.

Accordingly, in this section we consider the business case for investing in the means to acquire information about social networks. Can such knowledge be used to systematically improve audit yields through improved targeting? We also address the related questions of how the value of network information varies with the topological properties of the network, and with the assumed level of concern for social comparison. We begin by developing a theoretical framework for analyzing rigorously these questions, and then perform simulations of this framework to obtain numerical estimates.

4.1 Theoretical Framework

We consider the problem of a tax authority seeking to audit those taxpayers who have evaded most, conditional on observing (i) an income declaration d_i ; and (ii) some (potentially partial) information regarding the social network. Let the taxpayer’s income declaration be denoted as d_i . Using this notation, we may write evasion as $E_i = \theta(W_i) - \theta(d_i)$, thereby giving the income declaration as

$$d_i \equiv \hat{d}_i(\mathbf{G}) = \theta^{-1}(\theta(W_i) - E_i(W_i; \mathbf{G})). \quad (6)$$

The function $\hat{d}_i(W_i)$ in equation (6) gives the optimal disclosure d_i for a taxpayer with income W_i . Of relevance to our purpose, however, is the inverse function $W_i \equiv \hat{W}(d_i; \mathbf{G}) = \hat{d}_i^{-1}(d_i)$, which gives the true income W_i of a taxpayer who optimally declares an income d_i . On receipt of the taxpayer’s declaration d_i , a tax authority can use $\hat{W}(d_i; \cdot)$ to estimate the vector of true incomes \mathbf{W} . To illustrate, consider a tax authority that observes every aspect of the model except potentially the network. If such a tax authority perfectly observes the network, it can use $\hat{W}(d_i; \cdot)$ to correctly infer true income, i.e., $\hat{W}(d_i; \mathbf{G}) = W_i$. If, however, the tax authority does not perfectly observe the social network – instead observing some other (related) network $\mathbf{G}' \neq \mathbf{G}$ – it will obtain imperfect estimates of the W_i , given by $\hat{W}_i = \hat{W}(d_i; \mathbf{G}')$. The estimates \hat{W}_i may then be used to compute predicted evasion as $\hat{E}_i = \theta(\hat{W}_i) - \theta(d_i)$, which can be used to target audits towards those taxpayers with the highest \hat{E}_i . To the extent that the ordering of the \hat{W}_i does not match the ordering of the true W_i a failure to observe fully the social network leads to a suboptimal choice of audit targets.

To formalize this idea, we suppose that the tax authority observes only a subset of the links in the network. We define a probability $\kappa \in [0, 1]$ that the tax authority observes a given link in the social network. When $\kappa = 1$ the tax authority observes the social network perfectly; when $\kappa = 0.5$ on average half the links between taxpayers are known by the tax authority, and the other half are not; and when $\kappa = 0$ all links are unobserved. Hence, we obtain the adjacency matrix $\mathbf{G}(\kappa)$ from the “true” adjacency matrix \mathbf{G} by randomly deleting links (with probability $1 - \kappa$).⁷ Let $\mathcal{A}(p; \mathbf{G}(\kappa))$ be the set of taxpayers selected for audit (by virtue of having the highest values of \hat{E}_i), when a proportion p of taxpayers will be chosen for audit; and let $\mathfrak{R}_{\mathcal{A}}(p; \mathbf{G}(\kappa))$ be the revenue (in recovered taxes and fines) from auditing the taxpayers in \mathcal{A} . For a given p , audit revenue attains its theoretical maximum when the social network is observed perfectly, i.e., at $\mathfrak{R}_{\mathcal{A}}(p; \mathbf{G})$. At the other end of the information scale, any tax authority – no matter how information starved – can always choose to audit randomly. On average, a random audit strategy yields revenue $\mathfrak{R}_{RA}(p) = pf \sum_{j \in \mathcal{N}} E_j$.

To quantify the extent to which network information permits audits to be better-targeted,

⁷The only restriction we place on the random elimination of links is to ensure the requirement that each taxpayer’s reference group is non-empty. Accordingly, in cases where the random process that eliminates links chooses to eliminate all the links belonging to a taxpayer, one of the taxpayer’s links is randomly selected to be preserved.

we compare audit yields when audits are chosen utilizing network information to audit yields under a random audit rule. Our analysis centers on the statistic

$$\Psi(\kappa) \equiv \frac{\mathfrak{R}_{\mathcal{A}}(p; \mathbf{G}(\kappa)) - \mathfrak{R}_{RA}(p)}{\mathfrak{R}_{\mathcal{A}}(p; \mathbf{G}) - \mathfrak{R}_{RA}(p)} \times 100. \quad (7)$$

A value $\Psi(\kappa) = 1$ signifies that $\mathfrak{R}_{\mathcal{A}}(p; \mathbf{G}(\kappa))$ attains the full-information upper bound (in which audits are targeted perfectly). A value $\Psi(\kappa) = 0$ signifies that $\mathfrak{R}_{\mathcal{A}}(p; \mathbf{G}(\kappa))$ generates no improvement in revenues over random auditing. Intermediate values of $\Psi(\kappa)$ measure the improvement in revenue over random auditing, relative to that achieved by the full-information strategy.

4.2 Simulation

We now discuss how we simulate the model to obtain numerical estimates of the function $\Psi(\kappa)$.

4.2.1 The Social Network

To generate the social network \mathbf{G} , we follow the approach of network scientists, who utilize a class of network models, known as *generative models*, to investigate complex network formation (see, e.g., Pham *et al.*, 2016). In this modelling paradigm, complex networks are generated by means of the incremental addition of nodes and edges to a seed network over a long sequence of time-steps. Two processes governing the node/edge dynamics in generative models have been shown to generate features consistent with a multitude of social, biological, and technological networks (see, e.g., Redner, 1998; Adamic and Huberman, 2000; Jeong *et al.*, 2000; Ormerod and Roach, 2004; Capocci *et al.*, 2006). The first – the *node-degree* (or *preferential attachment*) process – makes the probability that a new taxpayer added to the network observes an existing taxpayer, i , a positive function of i 's degree (the number of taxpayers who already observe i). The second – the *node-fitness* process – makes the probability that a new taxpayer added to the network observes an existing taxpayer, i , a positive function of i 's *fitness* (an exogenous and time-invariant characteristic of node i).

At step s of the generative process consider a taxpayer i with degree \mathfrak{d}_{is} , and fitness $\eta_i > 0$. The separate node-degree and node-fitness processes are entwined in a single process by allowing the probability that said taxpayer i is observed by the taxpayer added at step s to be proportional to the product $\eta_i A(\mathfrak{d}_{is})$, where $A(\cdot)$ is an increasing function. Important

special cases of this approach include that of Barabási and Albert (1999), who assume η_i to be equal across taxpayers; and that of Bianconi and Barabási (2001), who assume $A(\vartheta) = \vartheta$. Recent research, however, suggests that social networks may be consistent with non-linear forms for $A(\cdot)$. In particular, the *sublinear* specification, $A(\vartheta) = \vartheta^\phi$, $\phi < 1$, finds empirical support (Backstrom *et al.*, 2006; Kunegis *et al.*, 2013; Pham *et al.*, 2016). Pham *et al.* (2016: 7) estimate $\phi = 0.43$ for the social network constituted by a sample of 46,000 Facebook wall-posts, and we adopt this estimate (we also investigate the systematic effects of varying this estimate of ϕ).

In allowing for a role for node-fitness in social network formation, we are able to account for the observation that, empirically, celebrity taxpayers are surely not drawn at random from the distribution of income, but rather belong systematically to the upper tail. TV and sports stars, whose consumption habits are widely reported, are also some of the richest members of society. To replicate this feature, we equate node-fitness with income: $\eta_i = W_i$. We specify the distribution function of W_i across taxpayers to satisfy a power law, consistent with a large body of empirical evidence (e.g., Coelho *et al.*, 2008).

In our implementation we generate networks of $N = 200$ taxpayers, starting from a seed network composed of two interlinked taxpayers. Each taxpayer incrementally added to the network is linked to members of the existing network according to the outcome of five random draws under the probability distribution $\eta_i \vartheta_{is}^\phi$ discussed above. Note, however, that these draws are with replacement, so a taxpayer may be linked multiple times to another. As the model of section 2 allows for only a single, albeit weighted, link between taxpayers, we use the frequency of links realized by the generative process to construct the comparison intensity weights. Specifically, let $\#_{ij} \in \mathbb{N}$ denote the number of times taxpayer i is linked with j by the generative process. If $\#_{ij} = 0$ then taxpayers i and j are not linked. If $\#_{ij} \geq 1$ then taxpayers i and j are linked, and the intensity of the link is given by

$$g_{ij} = \frac{\#_{ij}}{\sum_{k \in \mathcal{R}_i} \#_{ik}}.$$

Owing to its stochastic nature, any single iteration of the generative process may realize a network that is unrepresentative. To mitigate this concern we form \mathbf{G} as the average of multiple independent iterations of the generative process.⁸ With \mathbf{G} thus specified we

⁸That is, each element g_{ij} of \mathbf{G} is the mean of the realized comparison intensity weights across the independently generated networks. We do not average over a prescribed number of iterations, but rather implement a stopping rule that monitors the rate of convergence of the sample mean towards the true mean.

generate 14 $\mathbf{G}(\kappa)$ within the interval $\kappa = [0, 1]$. Each $\mathbf{G}(\kappa)$ is generated from \mathbf{G} by deleting links with probability $1 - \kappa$, and rescaling the resulting matrix such that $\sum_{j \in \mathcal{R}_i} g_{ij} = 1$ holds for the remaining links.

4.2.2 Model Functions and Parameters

Having now described the social network, we specify the remaining model functions and parameters. To make concrete the vector of predicted income, $\hat{\mathbf{W}}$, we specify the tax system as a linear income tax, $\theta(W_i) = \theta W_i$, where $\theta \in (0, 1)$. We may then write evasion as $E_i = \theta [W_i - d_i]$ and the legal disposable income level as $X(W_i) = [1 - \theta] W$. Next, we show that the vector $\hat{\mathbf{W}}$ takes the form of a weighted Bonacich centrality:

Lemma 1 *Under the conditions of Proposition 1, and with a linear income tax, the income of a taxpayer who declares income optimally according to (6) is given by*

$$\hat{\mathbf{W}}(\mathbf{d}; \mathbf{G}) = \mathbf{b}(\mathbf{V}, \theta, \gamma),$$

where

$$\begin{aligned} v_{ij} &= \frac{\zeta_i}{\xi_i} m_{ij}; & \xi_i &= [1 - \theta] [1 - p_i f] + \theta \{1 + [f - 2] p_i f\} > 0; \\ \gamma_{ij} &= \frac{\{1 + [f - 2] p_i f\} \theta a_i d_i + b_i [1 - p_i f]}{a_i \xi_i} + \frac{[1 - p_i f] R(\mathbf{X} - \theta [1 - p_i f] \mathbf{d})}{\xi_i}. \end{aligned}$$

Taxpayers are assumed to know the true average probability of audit, p , but do not know how the tax authority will select audit targets. Consistent with this idea, tax authorities are known to shroud their audit rules – the so-called “DIF score” in the case of the IRS – in great secrecy (see, e.g., Alm and McKee, 2004; Plumley and Steuerle, 2004; Hashimzade *et al.*, 2016). We set $\{p, f\}$ to be consistent with a level of evasion of ten percent, as is broadly consistent with the empirical evidence for developed countries cited in the Introduction. The level of evasion predicted by the model relates closely to the product pf , such that we are able to hold evasion at the ten percent level when, e.g., lowering f and raising p . The qualitative features of $\Psi(\kappa)$ that we shall report are unaffected by the chosen decomposition of pf , however.

We assume that the tax authority knows only average values of the parameters $\{a_i, b_i\}$, given by $\bar{a} = N^{-1} \sum_{j \in \mathcal{N}} a_j$ and $\bar{b} = N^{-1} \sum_{j \in \mathcal{N}} b_j$. This allows us to distinguish two separate cases. First, the baseline case we consider is when $\{a_i, b_i\}$ are the same for all taxpayers. In this

case $\{\bar{a}, \bar{b}\}$ coincides with the common $\{a, b\}$, so the tax authority is only ignorant of the network. Second, as a separate analysis we are then able to examine the effects of allowing unobserved heterogeneity in the $\{a_i, b_i\}$ to interact with partial observability of the social network. The baseline results are depicted for (common) utility parameters $\{a, b\} = \{2, 80\}$. Again, based on tests of the model predictions for a range of choices of $\{a, b\}$ consistent with an interior equilibrium, our qualitative findings are not sensitive to the particular choice of these two parameters.

4.2.3 Results

Implementing the methodology described above, our results for $\Psi(\kappa)$ under the baseline parameter specifications are shown in Figure 1. As would be expected, $\Psi(\kappa)$ is monotonically increasing in κ : more visibility of the network results in improved audit revenues relative to random auditing. It is apparent from the steepness of the left-side of the figure that there are significant returns from observing a little about the network (observing around 20 percent of all links) relative to knowing nothing. Thus, tax authorities that are in the infancy of their attempt to systematically construct social networks have an especially strong case to invest in this endeavour. The other remarkable feature of $\Psi(\kappa)$ at low values of κ is that, at values of κ very close to zero, it is actually counterproductive to seek to target audits on the basis of a very incomplete picture of the true social network ($\Psi(\kappa) < 0$). Rather, tax authorities in this situation are better-off simply choosing audit targets randomly. Only once approximately one percent of network links are observed does targeting based upon network information systematically improve upon random auditing. The strong initial returns to network visibility are seen to diminish on an interval of intermediate values of κ , before increasing strongly again as the tax authority moves from around 80 percent network visibility towards full visibility. Thus, in a way we have made precise, knowledge of the structure of social networks can be of value to tax authorities.

Figure 1 – see p. 26

It is of interest to understand how the value of network information is systematically affected by network structure. Figure 2 depicts the line of best fit for $\Psi(\kappa)$ as ϕ is varied around the benchmark value of 0.43. Specifically, we depict $\Psi(\kappa)$ for $\phi \in \{0.33, 0.43, 0.53\}$. Recall that ϕ regulates the importance of node-degree (preferential attachment) in the network formation

process: high values of ϕ produce networks with a highly concentrated distribution of links, implying the existence of a small number of extremely visible taxpayers. The left-side of the Figure indicates that, the greater the role of preferential attachment, the more deleterious are the results of attempting to utilize extremely incomplete information about the network. Intuitively, in cases of strong preferential attachment the true distribution of links will be highly concentrated. A successful audit strategy must therefore target a small number of key taxpayers. When network information is so scarce that the tax authority cannot observe the strong concentration of links, this leads to systematic misdirecting of audit resources. Above the neighborhood of $\kappa = 0$, however, preferential attachment reduces sensitivity to network visibility. As ϕ increases a growing proportion of taxpayers have the same few celebrity taxpayers in their reference group. It therefore becomes possible for the tax authority to identify these celebrity taxpayers, even when observing the network in a relatively limited way. In summary, the existence of celebrity taxpayers in social networks appears to present both threats and opportunities to tax authorities: it weakens the effectiveness auditing with very little information, but is helpful to tax authorities who can observe the network to a sufficient degree – and can thereby identify the celebrity taxpayers.

Figure 2 – see p. 27

The analysis so far has the feature that, as the parameters $\{a, b\}$ do not vary across taxpayers, it is only network information that is potentially unobserved by the tax authority. In practice, however, it seems certain that tax authorities also face unobserved preference heterogeneity. As a final analysis, therefore, we consider the robustness of our findings to the case in which unobserved preference heterogeneity coexists with partial observability of the network. To do this we generate the $\{a_i, b_i\}$ as realizations of the random processes $a \sim U(1.9, 2.1)$ and $b \sim U(76, 84)$. Note that the means of these two processes correspond to the $\{a, b\}$ in the baseline case, and that the bounds of the uniform distribution are set at ± 5 percent of these baseline values. The tax authority knows the mean of the processes generating the $\{a_i, b_i\}$, but not the realized values. In the presence of this unobserved preference heterogeneity we see (Figure 3) that, even when the social network is fully observed ($\kappa = 1$), the tax authority does not achieve the full-information revenue outcome. This effect aside, however, the qualitative shape of the function remains as in Figure 1, with the greatest benefits from network information appearing in the neighborhood of $\kappa = 0$ and $\kappa = 1$. Accordingly, we find

no evidence of important interaction effects between uncertainty over background preference parameters and uncertainty over the network.

Figure 3 – see p. 28

5 Conclusion

Tax evasion is estimated to cost governments of developed countries up to 20 percent of income tax revenues. We link the tax evasion decision with a large literature on the role in individual decision-making of social comparison.

Previous studies have restricted comparisons to be at the aggregate, rather than local, level. Moreover, the network structures that have been employed in these models possess few of the topological properties of observed social networks. In this paper we have sought to provide an analysis that addresses these issues. Taxpayers compare their consumption with others in their social network. In making social comparisons, each taxpayer makes “local” comparisons on their part of the social network. Engaging in tax evasion is a tool by which taxpayers can seek to increase their consumption relative to others. In this setting, we show that a linear-quadratic specification of utility yields a unique solution for optimal evasion corresponding to a weighted Bonacich centrality measure on a social network: by this measure, taxpayers that are more central in the social network evade more.

Our model provides a rich framework for understanding how information conveyed through a social network influences optimal evasion behavior. Although optimal evasion depends in quite a complex way on the underlying parameters, we are able to demonstrate how an increase in income of one taxpayer in the social network leads other taxpayers to optimally increase their evasion, while an increase in the probability of tax authority enforcement induces other taxpayers to reduce their evasion. The size of these effects is heterogeneous across taxpayers owing to heterogeneity in how taxpayers are linked in the social network.

Given that tax authorities are now investing in technology that seeks to construct social networks, we show that network information can allow a tax authority to better predict the likely revenue benefits from conducting an audit of a particular taxpayer. We obtain numerical estimates of this effect using a social network that permits, in particular, the

existence highly-observed “celebrity” taxpayers belonging systematically to the upper tail of the income distribution. Our results point to an important role for network effects. In particular, for a tax authority that is largely ignorant of the social network, we document strong initial revenue gains from acquiring relatively small amounts of network information. We also have a cautionary message, however, that attempting to target audits on the basis of extremely limited network information may be counterproductive, especially if the true distribution of links is highly concentrated. In this situation a random audit strategy is more gainful.

We finish with some possible avenues for future research. First, it would also be of interest to introduce dynamic features to the model that relate behavior today to past evasion decisions and audit outcomes. Second, while we have focused on tax evasion, early empirical work (Alstadsæter *et al.*, 2018) suggests the relevance of a similar modelling approach to tax avoidance behavior, or indeed criminal activity more generally. While these extensions must await a dedicated treatment, we hope our contribution at least clarifies the role of social comparison in driving tax evasion behavior on a social network.

References

- Adamic, L.A. and Huberman, B.A. (2000). “Power-law distribution of the World Wide Web”, *Science* 287(5461), 2115.
- Alm, J., Bloomquist, K.M., and McKee, M. (2017). “When you know your neighbour pays taxes: Information, peer effects and tax compliance”, *Fiscal Studies* 38(4), pp. 587–613.
- Alm, J. and McKee, M. (2004). “Tax compliance as a coordination game”, *Journal of Economic Behavior & Organization* 54(3), pp. 297–312.
- Alm, J., Jackson, B.R., and McKee, M. (2009). “Getting the word out: Enforcement information dissemination and compliance behavior”, *Journal of Public Economics* 93(3-4), pp. 392–402.
- Alm, J. and Yunus, M. (2009). “Spatiality and persistence in U.S. individual income tax compliance”, *National Tax Journal* 62(1), pp. 101–124.
- Alstadsæter, A., Kopczuk, W., and Telle, K. (2018). “Social networks and tax avoidance: Evidence from a well-defined Norwegian tax shelter”. Paper presented at the 74th Annual Congress of the International Institute of Public Finance, August, University of Tampere.
- Andrei, A.L., Comer, K., and Koelher, M. (2014). “An agent-based model of network effects on tax compliance and evasion”, *Journal of Economic Psychology* 40(1), pp. 119–133.

- Backstrom, L., Huttenlocher, D., Kleinberg, K., and Lan, X. (2006). “Group formation in large social networks: membership, growth, and evolution”. In *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 44-54, New York: Association for Computing Machinery.
- Baldwin, R. and McKenna, A. (2014) “Well connected”, *Taxation* 174 (4467).
- Ballester, C., Calvó-Armengol, A., and Zenou, Y. (2006). “Who’s who in networks. Wanted: the key player”, *Econometrica* 74(5), pp. 1403–1417.
- Barabási, A.-L. and Albert, R. (1999). “Emergence of scaling in random networks”, *Science* 286 (5439), pp. 509–512.
- Becker, G., Murphy, K., and Werning, I. (2005). “The equilibrium distribution of income and the market for status”, *Journal of Political Economy* 113(2), pp. 282–310.
- Bernasconi, M., Levaggi, R., and Menoncin, F. (2016). “Dynamic tax evasion with habit formation”, Working Paper No. 31, Department of Economics, Ca’ Foscari University of Venice.
- Bianconi, G. and Barabási, A.-L. (2001). “Competition and multiscaling in evolving networks”, *Europhysics Letters* 54(4), pp. 436–442.
- Bloomquist, M.K. (2011). “Tax compliance as an evolutionary coordination game: An agent-based approach”, *Public Finance Review* 39(1), pp. 25–49.
- Bonacich, P. (1987). “Power and centrality: A family of measures”, *American Journal of Sociology* 92(5), pp. 1170–1182.
- Boskin, M. and Sheshinski, E. (1978). “Optimal redistributive taxation when individual welfare depends upon relative income”, *Quarterly Journal of Economics* 92(4), pp. 589–601.
- Bramoullé, Y. and Kranton, R. (2007). “Public goods and networks”, *Journal of Economic Theory* 135(1), pp. 478–494.
- Buehn, A. and Schneider, F. (2016). “Size and development of tax evasion in 38 OECD countries: What do we (not) know?”, *Journal of Economics and Political Economy* 3(1), DOI: 10.1453/jepe.v3i1.634
- Calvó-Armengol, A. and Jackson, M.O. (2004). “The effects of social networks on employment and inequality”, *American Economic Review* 94(3), pp. 426–454.
- Capocci, A., Servedio, V.D.P., Colaiori, F., Buriol, L.S., Donato, D., Leonardi, S., and Caldarelli, G. (2006). “Preferential attachment in the growth of social networks: The internet encyclopedia Wikipedia”, *Physical Review E* 74(3). DOI: 10.1103/PhysRevE.74.036116
- Cebula, R.J. and Feige, E.L. (2012). “America’s unreported economy: measuring the size, growth and determinants of income tax evasion in the U.S.”, *Crime, Law and Social Change* 57(3), pp. 265–285.
- Clark, A.E., Frijters, P., and Shields, M.A. (2008). “Relative income, happiness, and utility: An explanation for the Easterlin paradox and other puzzles”, *Journal of Economic Literature* 46(1), pp. 95–144.

- Clark, A.E. and Senik, C. (2010). “Who compares to whom? The anatomy of income comparisons in Europe”, *Economic Journal* 120(544), pp. 573–594.
- Coelho, R., Richmond, P., Barry, J., and Hutzler, S. (2008). “Double power laws in income and wealth distributions”, *Physica A* 387(15), pp. 3847–3851.
- Davis, J.S., Hecht, G., and Perkins, J.D. (2003). “Social behaviors, enforcement, and tax compliance dynamics”, *Accounting Review* 78(1), pp. 39–69.
- De Juan, A., Lasheras, M.A., and Mayo, R. (1994). “Voluntary tax compliant behavior of Spanish income tax payers”, *Public Finance / Finances Publiques* 49(S), pp. 90–105.
- Dell’Anno, R. (2009). “Tax evasion, tax morale and policy maker’s effectiveness”, *Journal of Socio-Economics* 38(6), pp. 988–997.
- Edlin, A.S. and Shannon, C. (1998). “Strict monotonicity in comparative statics”, *Journal of Economic Theory* 81(1), pp. 201–219.
- Ferrer-i-Carbonell, A. (2005). “Income and well-being: an empirical analysis of the comparison income effect”, *Journal of Public Economics* 89(5-6), pp. 997–1019.
- Fortin, B., Lacroix, G., and Villeval, M.-C. (2007). “Tax evasion and social interactions”, *Journal of Public Economics* 91(11-12), pp. 2089–2112.
- Foster, J.G., Foster, D.V., Grassberger, P., and Paczuski, M. (2010). “Edge direction and the structure of networks”, *PNAS* 107(24), pp. 10815–10820.
- Frey, B.S. and Stutzer, A. (2002). *Happiness and Economics*, Princeton: Princeton University Press.
- Glaeser, E., Sacerdote, B., and Scheinkman, J. (1996). “Crime and social interactions”, *Quarterly Journal of Economics* 111(2), pp. 507–548.
- Goerke, L. (2013). “Relative consumption and tax evasion”, *Journal of Economic Behavior & Organization* 87(1), pp. 52–65.
- Gordon, J.P.P. (1989). “Individual morality and reputation costs as deterrents to tax evasion”, *European Economic Review* 33(4), pp. 797–805.
- Harbaugh, R. (1996). “Falling behind the Joneses: Relative consumption and the growth-savings paradox”, *Economics Letters* 53(3), pp. 297–304.
- Hashimzade, N., Myles, G.D., Page, F., and Rablen, M.D. (2014). “Social networks and occupational choice: The endogenous formation of attitudes and beliefs about tax compliance”, *Journal of Economic Psychology* 40(1), pp. 134–146.
- Hashimzade, N., Myles, G.D., and Rablen, M.D. (2016). “Predictive analytics and the targeting of audits”, *Journal of Economic Behavior & Organization* 124(1), pp. 130–145.
- H.M. Revenue and Customs (2016). *Measuring Tax Gaps 2016 Edition: Tax Gap Estimates for 2014-15*. London: H.M. Revenue and Customs.
- Hokamp, S. (2014). “Dynamics of tax evasion with back auditing, social norm updating, and public goods provision – an agent-based simulation”, *Journal of Economic Psychology* 40(1), pp. 187–199.

- Hokamp, S. and Pickhardt, M. (2010). “Income tax evasion in a society of heterogeneous agents – evidence from an agent-based model”, *International Economic Journal* 24(4), pp. 541–553.
- Hopkins, E. and Kornienko, T. (2004). “Running to keep in the same place: Consumer choice as a game of status”, *American Economic Review* 94(4), pp. 1085–1107.
- Ioannides, Y. (2012). *From Neighborhoods to Nations: The Economics of Social Interactions*, Princeton: Princeton University Press.
- Jackson, M.O., Rogers, B.W., and Zenou, Y. (2017). “The economic consequences of social-network structure”, *Journal of Economic Literature* 55(1), pp. 49–95.
- Jackson, M.O. and Zenou, Y. (2015). Games on networks. In P. Young and S. Zamir (Eds.), *Handbook of Game Theory*, Vol. 4, pp. 34–61, Amsterdam: Elsevier.
- Jeong, H., Tombor, B., Albert, R., Oltvai, Z., and Barabási, A. (2000). “The large-scale organization of metabolic networks”, *Nature* 407, pp. 651–654.
- Kim, Y. (2003). “Income distribution and equilibrium multiplicity in a stigma-based model of tax evasion”, *Journal of Public Economics* 87(7-8), pp. 1591–1616.
- Korobow, A., Johnson, C., and Axtell, R. (2007). “An agent-based model of tax compliance with social networks”, *National Tax Journal* 60(3), pp. 589–610.
- Kunegis, J., Blattner, M., and Moser, C. (2013). “Preferential attachment in online networks: measurement and explanations”. In *Proceedings of the 5th Annual ACM Web Science Conference*, pp. 205–214, New York: Association for Computing Machinery.
- Layard, R. (1980). “Human satisfactions and public policy”, *Economic Journal* 90(360), pp. 737–750.
- Layard, R. (2005). *Happiness. Lessons from a New Science*, London: Allen Lane.
- Ljungqvist, L. and Uhlig, H. (2000). “Tax policy and aggregate demand management under catching up with the Joneses”, *American Economic Review* 90(3), pp. 356–366.
- Luttmer, E.F.P. (2005). “Neighbors as negatives: Relative earnings and well-being”, *Quarterly Journal of Economics* 120(3), pp. 963–1002.
- Mittone, L. and Patelli, P. (2000). “Imitative behaviour in tax evasion”. In F. Luna and B. Stefansson (Eds.), *Economic Modelling with Swarm*, Ch. 5, Amsterdam: Kluwer.
- Mujcic, R. and Frijters, P. (2013). “Economic choices and status: measuring preferences for income rank”, *Oxford Economic Papers* 65(1), pp. 47–73.
- Myles, G.D. and Naylor, R.A. (1996). “A model of tax evasion with group conformity and social customs”, *European Journal of Political Economy* 12(1), pp. 49–66.
- Neumark, D. and Postlewaite, A. (1998). “Relative income concerns and the rise in married women’s employment”, *Journal of Public Economics* 70(1), pp. 157–183.
- Ormerod, P. and Roach, A.P. (2004). “The medieval inquisition: scale-free networks and the suppression of heresy”, *Physica A: Statistical Mechanics and its Applications* 339(3), pp. 645–652.

- Panadés, J. (2004). “Tax evasion and relative tax contribution”, *Public Finance Review* 32(2), pp. 183–195.
- Pham, T., Sheridan, P., and Shimodaira, H. (2016). “Joint estimation of preferential attachment and node fitness in growing complex networks” *Scientific Reports* 6: 32558. DOI: 10.1038/srep32558
- Plumley, A. (1996). “The determinants of individual income tax compliance” Internal Revenue Service Publication 1916 (Rev. 11-96), IRS: Washington, DC.
- Plumley, A.H. and Steuerle, C.E. (2004). “Ultimate objectives for the IRS: Balancing revenue and service”. In H.J. Aaron and J. Slemrod (eds.) *The Crisis in Tax Administration*, pp. 311–346, Washington, DC: Brookings Institution Press.
- Postlewaite, A. (1998). “The social basis of interdependent preferences”, *European Economic Review* 42(3-5), pp. 779–800.
- Rablen, M.D. (2008). “Relativity, rank and the utility of income”, *Economic Journal* 118(528), pp. 801–821.
- Ratto, M., Thomas, R., and Ulph, D. (2013). “The indirect effects of auditing taxpayers”, *Public Finance Review* 41(3), pp. 317–333.
- Rayo, L. and Becker, G.S. (2007). “Evolutionary efficiency and happiness”, *Journal of Political Economy* 115(2), pp. 302–337.
- Redner, S. (1998). “How popular is your paper? An empirical study of the citation distribution”, *European Physical Journal B – Condensed Matter and Complex Systems* 4, pp. 131–134.
- Rigney, P. (2016). *The All Seeing Eye – An HMRC Success Story?* London: Institute of Financial Accountants.
- Samuelson, L. (2004). “Information-based relative consumption effects”, *Econometrica* 72(1), pp. 93–118.
- Sen, A. (1983). “Poor, relatively speaking”, *Oxford Economic Papers* 35(2), pp. 153–169.
- Stark, O. and Taylor, J. (1991). “Migration incentives, migration types: the role of relative deprivation”, *Economic Journal* 101(408), pp. 1163–1178.
- Suter, L. (2017). *Taxman Unleashes Snooper Computer: What Information Does it Have on You?*, 7th January, London: Telegraph Media Group.
- Szell, M. and Thurner, S. (2010). “Measuring social dynamics in a massive multiplayer online game”, *Social Networks* 32(4), pp. 313–329.
- Traxler, C. (2010). “Social norms and conditional cooperative taxpayers”, *European Journal of Political Economy* 26(1), pp. 89–103.
- Tremblay, C.H. and Tremblay, V.J. (2010). “The neglect of monotone comparative statics methods”, *Journal of Economic Education* 41(2), pp. 177–193.

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

Yitzhaki, S. (1974). “A note on ‘Income tax evasion: A theoretical analysis’”, *Journal of Public Economics* 3(2), pp. 201–202.

Zaklan, G., Lima, F.W.S., and Westerhoff, F. (2008). “Controlling tax evasion fluctuations”, *Physica A: Statistical Mechanics and its Applications* 387(23), pp. 5857–5861.

Appendix

Proof of Proposition 1. Under linear-quadratic utility equation (3) can be solved to give optimal evasion at an interior solution as

$$E_i = \frac{1 - p_i f}{a_i \zeta_i} \{b_i - a_i [X_i - R_i]\}, \quad (\text{A.1})$$

where $\zeta_i > 0$ is defined in the Proposition. Marginal utility, $b_i - a_i [X_i - R_i]$, is positive by the assumed restrictions on $\{a_i, b_i\}$. Using (4), optimal evasion in (A.1) is written in full as

$$E_i = \frac{1 - p_i f}{a_i \zeta_i} \{b_i - a_i [R(\mathbf{X}) + [1 - p_i f] \mathbf{g}_i \mathbf{E}]\}, \quad (\text{A.2})$$

where necessarily $1 - p_i f > 0$ at an interior optimum. Then the set of N equations defined by (A.2) for taxpayers $i \in \mathcal{N}$ can be written in matrix form as $\mathbf{E} = \boldsymbol{\alpha} + \mathbf{M}\mathbf{E}$ where the elements of $\{\boldsymbol{\alpha}, \mathbf{1}, \mathbf{M}\}$ are as in Proposition 1. It follows that $[\mathbf{I} - \mathbf{M}]\mathbf{E} = \boldsymbol{\alpha}$, so $\mathbf{E} = [\mathbf{I} - \mathbf{M}]^{-1} \boldsymbol{\alpha} \equiv \mathbf{b}(\mathbf{M}, \mathbf{1}, \boldsymbol{\alpha})$. ■

Proof of Proposition 2. We have

$$\begin{aligned} \frac{\partial \mathbf{E}}{\partial p_j} &= \frac{\partial [\mathbf{I} - \mathbf{M}]^{-1}}{\partial p_j} \boldsymbol{\alpha} + [\mathbf{I} - \mathbf{M}]^{-1} \frac{\partial \boldsymbol{\alpha}}{\partial p_j}; \\ &= [\mathbf{I} - \mathbf{M}]^{-1} \frac{\partial \mathbf{M}}{\partial p_j} [\mathbf{I} - \mathbf{M}]^{-1} \boldsymbol{\alpha} + [\mathbf{I} - \mathbf{M}]^{-1} \frac{\partial \boldsymbol{\alpha}}{\partial p_j}; \\ &= [\mathbf{I} - \mathbf{M}]^{-1} \left[\frac{\partial \mathbf{M}}{\partial p_j} [\mathbf{I} - \mathbf{M}]^{-1} \boldsymbol{\alpha} + \frac{\partial \boldsymbol{\alpha}}{\partial p_j} \right]; \\ &= [\mathbf{I} - \mathbf{M}]^{-1} \left[\frac{\partial \mathbf{M}}{\partial p_j} \mathbf{E} + \frac{\partial \boldsymbol{\alpha}}{\partial p_j} \right]; \\ &= \mathbf{b} \left(\mathbf{M}, \mathbf{1}, \frac{\partial \mathbf{M}}{\partial p_j} \mathbf{E} + \frac{\partial \boldsymbol{\alpha}}{\partial p_j} \right); \\ \frac{\partial \mathbf{E}}{\partial W_j} &= \frac{\partial \mathbf{E}}{\partial X_j} = [\mathbf{I} - \mathbf{M}]^{-1} \frac{\partial \boldsymbol{\alpha}}{\partial X_j} = \mathbf{b} \left(\mathbf{M}, \mathbf{1}, \frac{\partial \boldsymbol{\alpha}}{\partial X_j} \right); \end{aligned}$$

from which the Proposition follows. ■

Proof of Lemma 1. Substituting $X(W_i) = [1 - \theta] W_i$ and $E_i = \theta [W_i - d_i]$ into (A.1) and rearranging for W_i gives

$$\begin{aligned} W_i &= \frac{\{1 + [f - 2] p_i f\} \theta a_i d_i + b_i [1 - p_i f] + a_i [1 - p_i f] R_i}{a_i \xi_i}; \\ \xi_i &= [1 - \theta] [1 - p_i f] + \theta \{1 + [f - 2] p_i f\}. \end{aligned} \quad (\text{A.3})$$

Noting that the second order condition for (A.1) to define a maximum is $-\theta^2 a_i \{1 + [f - 2] p_i f\} < 0$, it follows that $\xi_i > 0$. Using (4), (A.3) is written in full as

$$W_i = \frac{\{1 + [f - 2] p_i f\} \theta a_i d_i + b_i [1 - p_i f]}{a_i \xi_i} + \frac{a_i [1 - p_i f] \{R(\mathbf{X} - \theta [1 - p_i f] \mathbf{d}) + [1 - p_i f] [1 - p_i f] \mathbf{g}_i \mathbf{W}\}}{a_i \xi_i}. \quad (\text{A.4})$$

Then the set of N equations defined by (A.4) for taxpayers $i \in \mathcal{N}$ can be written in matrix form as $\mathbf{W} = \boldsymbol{\gamma} + \theta \mathbf{V} \mathbf{W}$ where the elements of $\{\boldsymbol{\gamma}, \theta, \mathbf{V}\}$ are as in Proposition 1. Hence $\mathbf{W} = [\mathbf{I} - \theta \mathbf{V}]^{-1} \boldsymbol{\gamma} = \mathbf{b}(\mathbf{V}, \theta, \boldsymbol{\gamma})$. ■

Figures

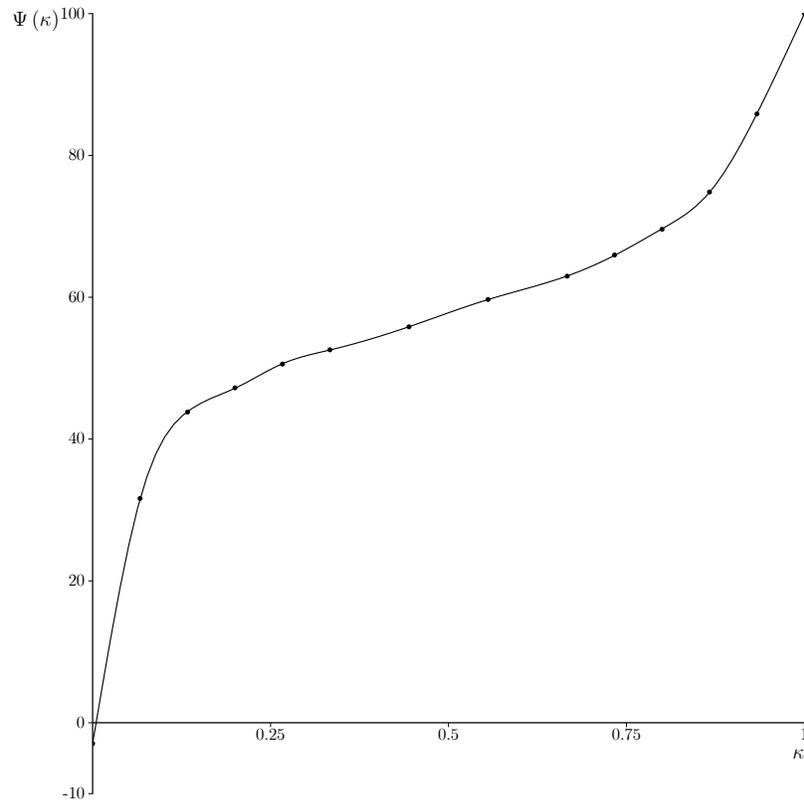


Figure 1: Revenue effects of network information.

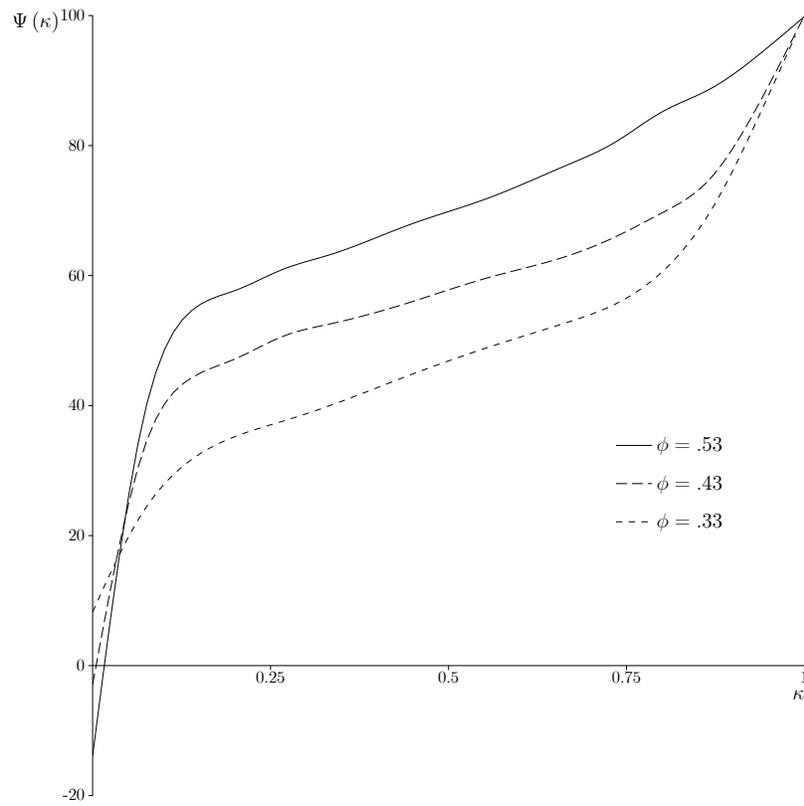


Figure 2: The role of network structure.

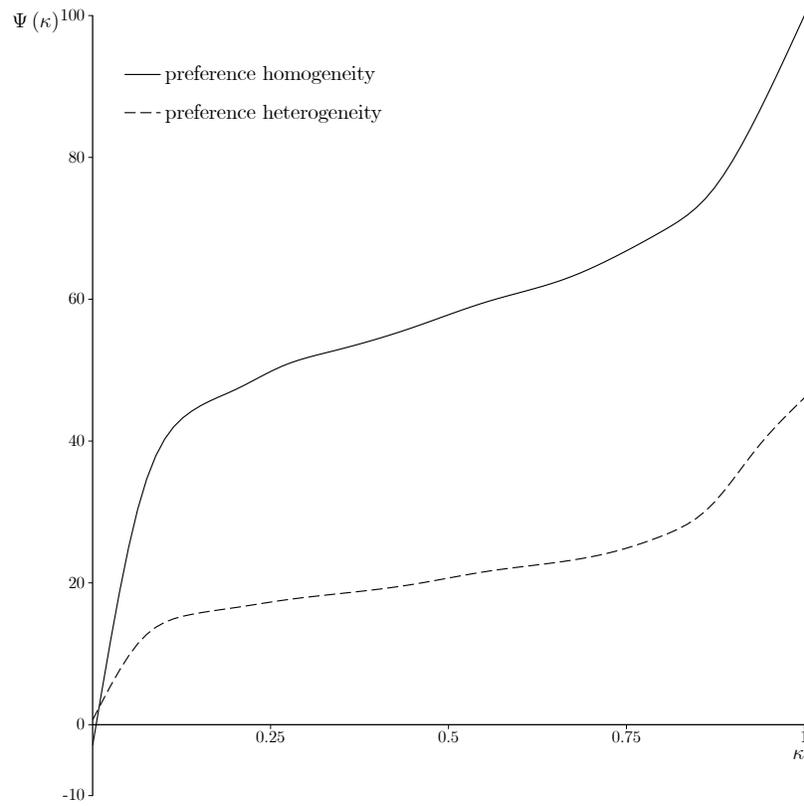


Figure 3: Revenue effects under unobserved preference heterogeneity.