The Tax Evasion Social Multiplier: Evidence from Italy

Roberto Galbiati* Giulio Zanella†

First Draft: September 2008
This Draft: January 2012

Abstract

We estimate social externalities of tax evasion in a model where congestion of the auditing resources of local tax authorities generates a social multiplier. Identification is based on a contrast of the variance of tax evasion at different levels of aggregation. We use a unique data set that contains audits of about 80,000 small businesses and professionals in Italy and also provides an exact measure of reference groups in our model. We find a social multiplier of about 3, which means that the equilibrium response to a shock that induces an exogenous variation in mean concealed income is about 3 times the initial average response. This is a short-run effect that persists to the extent that auditing resources are not adjusted to internalize the congestion externality.

JEL codes: H26, C31, Z13, Z19.
Keywords: social interactions, neighborhood effects, social multiplier, tax evasion, tax compliance, illegal behavior, excess variance

*CNRS EconomiX and Sciences Po, Paris, France. E-mail: galbiatir@gmail.com

†Corresponding author. University of Bologna, Italy, Department of Economics, Piazza Scaravilli 2, 40126 Bologna, Italy. E-mail: giulio.zanella@unibo.it

Theft—whether from the state, from a fellow citizen or from a looted Jewish store—was so widespread that in the eyes of many people it ceased to be a crime.

—Tony Judt, *Postwar*.[1]

1 Introduction

In this paper, we investigate the social determinants of tax compliance and tax evasion. Like most other kinds of illegal behavior, tax evasion exhibits large variance across geographic units with relatively similar fundamentals, such as similar countries or areas within a country.[2] In the benchmark model of Allingham and Sandmo (1972) such fundamentals are the parameters characterizing preferences (degree of risk aversion), the tax system (tax rates) and the enforcement system (probability of detection and sanctions).[3] In this paper we use detailed audit data from Italy to show that the observed large variance of tax evasion in spite of similar fundamentals reflects, to some extent, social externalities in underreporting income. We emphasize a particular source of such externalities: tax enforcement congestion.

Generally speaking, although large residual variance in illegal behavior—and many other types of socioeconomic phenomena—may be due to mere unobserved heterogeneity, social scientists view such variance increasingly as a telltale sign of interdependencies between individual decision makers. Glaeser, Sacerdote, and Scheinkman (1996) pioneered this approach by showing formally how positive covariance between individual decisions to engage in crime generates a multiplier effect that amplifies—both in time and across space—relatively small differences in fundamentals. The reason

[1] Judt (2005), 37.

[2] In Italy, for instance, the picture varies greatly from region to region and also across provinces within regions, even though Italian regions are quite homogeneous units. Pisani and Polito (2006) estimate that between 1998 and 2002 the ratio of concealed to reported income from productive activities across Italian regions ranged from 13% in Lombardy and 22% in Emilia Romagna and Veneto to 66% in Sicily and 94% in Calabria. The variance across provinces within regions is noticeable. Two extreme examples are Lombardy, where the ratio ranges from 5% in the province of Milan to 34% in the neighboring province of Lodi, and Calabria, where it ranges from 53% in the province of Reggio Calabria to 184% in the province of Vibo Valentia.

[3] An extension of this model allows for "tax morale", an intrinsic motivation inducing people to abide by their tax obligations—i.e., an additional preference parameter. See, for example, the theoretical analysis of Gordon (1989) and the empirical study by Frey and Feld (2002). Andreoni, Erard, and Feinstein (1998) and Slemrod (2007) offer excellent surveys of theory and evidence on tax compliance and tax evasion. Sandmo (2005) offers a retrospective discussion of the shortcomings and of the unexploited potential of the Allingham—Sandmo model.

is that, in the presence of positive complementarities between individual choices, any shock affects individual behavior directly via private incentives and indirectly via the behavior of other individuals. The ratio between the equilibrium aggregate response to the shock and the sum of the direct, initial individual responses is the *social multiplier*.[4] For the most, the literature interprets these externalities as having been generated by sociological forces embedded in individual preferences—what Manski (2000) classifies as "preference" interactions. For example, the seminal paper of Allingham and Sandmo (1972) includes an extended version of the basic model that features a social stigma effect; Gordon (1989) introduces the idea of tax morale sustained by peer pressure in a model of tax evasion; Cowell (1990, chapter 6) analyzes equilibrium tax evasion when preferences depend on the average evasion of other taxpayers; and Myles and Naylor (1996) analyze an optimal audit policy for an independent revenue service when there is a social custom that rewards honest tax paying. In this paper we emphasize a potentially important source of social complementarity that is more technological in nature and thus more amenable to policy: enforcement congestion. The idea is that the probability of apprehension and punishment decreases if more people behave illegally while the enforcer's available resources are fixed. Manski (2000) classifies these as "constraint" interactions. The importance of constraint interactions for illegal behavior is discussed by Ehrlich (1973), Sah (1991), and, more recently, Ferrer (2010). In the context of tax evasion, this externality is implicit in models where taxpayers and the tax authority interact strategically and the latter is subject to a budget constraint in its auditing activity (Sánchez and Sobel, 1993; Bassetto and Phelan, 2008).

We allow for tax enforcement congestion in a simple model where taxpayers belong to local tax jurisdictions and decide how much to report to a local tax authority, which performs audits subject to a budget constraint. If the individual probability of an audit is decreasing in individual reported income and the local budget constraint cannot be relaxed promptly, then a social effect arises: when some taxpayers report less income, the probability of other taxpayers in that jurisdiction being audited decreases. Hence these other taxpayers will also report less income. In the social interactions literature, the group that influences the behavior of an individual is called that individual's *reference group*. Therefore, local tax jurisdictions are natural reference groups in the tax evasion model we describe here. The equilibrium maps into the popular linear-in-means model frequently employed in

[4] We believe the term *social multiplier* was first used in the sense in which it is now commonly used in the social interactions literature by Schlicht (1981).

empirical analyses of social interactions (Manski, 1993). Thus our estimates admit a structural interpretation.

Identification exploits the "variance contrasts" method developed by Graham (2008), who extends the framework of Glaeser, Sacerdote, and Scheinkman (1996). The key idea is that the social multiplier can be identified by comparing the within-group and between-group variance of individual behavior (i.e., the same variance at different levels of aggregation) provided at least one group-level exogenous characteristic affects the within-group variance but does not directly affect the between-group variance. As in Graham (2008), the typically larger dispersion of individual heterogeneity (and so of tax evasion) in small reference groups provides such an identifying restriction. We show that the social multiplier can be identified in this way while retaining a structural interpretation in terms of only endogenous social effects.

Our empirical analysis employs a unique, cross-sectional data set of tax audits of self-employed workers in Italy (i.e., of small businesses and professionals). The audits we observe were performed by local branches of the national fiscal authority. These branches are responsible for tax enforcement within local tax jurisdictions. We thus observe the exact measure of reference groups in our model—a unique feature among nonexperimental studies of social interactions. We find a social multiplier of about 3, which means that the equilibrium aggregate response to a shock that affects concealed income is about 3 times the initial average response. This result has noteworthy policy implications. We mention two of them here, postponing a thorough discussion until the end of the paper. First, reducing tax evasion may be easier than generally supposed, because the social multiplier amplifies the impact of stricter enforcement. In other words, governments can reduce tax evasion at a fraction of the cost needed to directly induce each taxpayer to report more honestly. Conversely, looser enforcement reduces tax revenues more than when multiplier effects are absent. Second, if individual incentives change in favor of underreporting then the government should promptly adjust its auditing resources in order to internalize the congestion externality and prevent an outbreak of tax evasion.

The paper is organized as follows. Section 2 presents the model, and Section 3 describes the data set and its institutional background. The formal econometric framework is presented in Section 4 along with the identification strategy. Results are reported in Section 5, and Section 6 concludes. A Supplemental web appendix (Galbiati and Zanella, 2012) contains all derivations, some extensions, and additional material.

2 Model

2.1 Setup and equilibrium

Our empirical analysis will exploit data on income reports by self-employed workers. Because the personal income of these workers is not subject to third-party report, the Allingham–Sandmo model is particularly apt for interpreting these data. Consider a population of N taxpayers, indexed by $i = 1, ..., N$, distributed across G tax jurisdictions, indexed by $g = 1, ..., G$ and of size n_g. Local tax authorities are in charge of tax enforcement in each group g, and they receive from the central government $a_g \leq n_g$ tokens that can be used to audit taxpayers. The cost of an audit is one token. The taxable income of taxpayer i is private information and is denoted y_i. The taxpayer reports an amount y_i^R, which is public information, to the local tax authority and pays a tax at an exogenous, individual-specific flat rate t_i on this amount. Taxpayer i in jurisdiction g is audited with probability p_{ig}. We assume that an audit enables the tax authority to observe true taxable income. If $y_i^R = y_i$ then nothing happens, but if $y_i^R < y_i$ then the taxpayer must pay the full tax bill as well as a proportional fine at rate f on the evaded tax.[5] We assume that there are no rebates when a taxpayer overreports income (i.e., when $y_i^R > y_i$). Hence the taxpayer will never overreport and so we can ignore that case in what follows.

We assume that the taxpayer is risk neutral.[6] Therefore, we follow Scotchmer (1987) in assuming that the goal of the taxpayer is to minimize the expected tax bill:

$$\min_{y_i^R} (1 - p_{ig}) t_i y_i^R + p_{ig} t_i (y_i + f(y_i - y_i^R)). \tag{1}$$

We do not model explicitly the determination of audit probabilities. Rather, we posit a linear specification that captures in a simple way the externality arising from the tax authority budget constraint when auditing resources are given:

[5] In assuming that the fine is proportional to the evaded tax (and not to undeclared income) we adopt the Yitzhaki (1974) variant of the Allingham–Sandmo model, although which version is adopted is not important for our purposes.The Yitzhaki variant simplifies the model because it rules out the substitution effect when the tax rate changes.

[6] This assumption is appealing in the context of this paper because our data set consists of a sample of entrepreneurs.

$$p_{ig} = \frac{a_g}{n_g} + \alpha_0 \Pr(y_i^R < y_i \mid \mathbf{x}_i) - \frac{\alpha_1}{n_g - 1} \sum_{j=1, j\neq i}^{n_g} \Pr(y_j^R < y_j \mid \mathbf{x}_j). \tag{2}$$

Here α_0 and α_1 are positive parameters, and $\mathbf{x}$ is a vector of individual characteristics observable to the tax authority. It is understood, as in any linear probability model, that p_{ig} is defined to be 1 (resp., 0) if the right-hand side of (2) exceeds unity (resp., is negative). The first term on the RHS of (2) is the baseline probability of an audit: if the tax authority audited taxpayers at random, then a_g/n_g is the probability that any taxpayer in group g is audited. The second term reflects the idea that, the higher the probability (as estimated by the tax authority on the basis of $\mathbf{x}$) of taxpayer i concealing income, the higher the likelihood of an audit. The last term reflects this same idea for the other taxpayers in the jurisdiction and thus represents a local externality: if the other taxpayers in the group become more likely cheaters in the eyes of the tax authority and if the number of individuals that can be audited (i.e., a_g) is given, then those other taxpayers will force the local authority to shift auditing resources away from taxpayer i. In other words, when the resource constraint of the local auditor is binding and taxpayer i conceals more income it becomes less likely that other taxpayers in the jurisdiction will be audited. We assume that the congestion effect is uniform, whence the magnitude of the externality is normalized by group size after removing individual i (i.e., $n_g - 1$).

Equation (2) collapses the dynamic process described by Sah (1991) into a static setting. Sah describes a dynamic model in which potential criminals update the perceived probability of detection after observing how many people in their social neighborhood engage in crime and how many are apprehended. In Sah's model, the actual apprehension rate decreases with the number engaging in crime because resources cannot adjust immediately. This lower apprehension rate feeds back into the perceived probability of apprehension, which decreases as a consequence. Therefore, our static formulation is consistent with the cross-sectional implications of dynamic models of enforcement congestion.

We impose further linearities on our framework by assuming that the two probabilities on the RHS of equation (2) are themselves governed by a linear probability model. That is: for every individual i and for parameters $\beta_0 > 0$ and $\boldsymbol{\beta}_1$,

$$\Pr(y_i^R < y_i \mid \mathbf{x}_i) = 1 - \Pr(y_i \leq y_i^R \mid \mathbf{x}_i) = 1 - \beta_0 y_i^R - \boldsymbol{\beta}_1 \mathbf{x}_i, \tag{3}$$

where we retain our previous understanding of this function's behavior. The linearities introduced by equations (2) and (3) are somewhat ad hoc, but they offer the important advantage of allowing us to derive clearly the linear-in-means model of social interactions from a microfounded model of taxpayer behavior. This derivation is illustrated next.

At an interior optimum, the first-order necessary conditions for a minimum[7] in the taxpayer's problem exactly balance the expected marginal benefit and cost of underreporting income, *net* of the marginal effect on the probability of an audit.[8] That is:

$$t_i(1 - p_{ig}) = t_i p_{ig} f - \frac{\partial p_{ig}}{\partial y_i^R} t_i(1+f)\left(y_i - y_i^R\right). \tag{4}$$

Under the large group approximation

$$(n_g - 1)^{-1} \sum_{j=1, j\neq i}^{n_g} y_j^R \simeq n_g^{-1} \sum_{i=1}^{n_g} y_i^R = \mathbb{E}(y_i^R|g), \tag{5}$$

if we substitute expression (2) into the taxpayer's first-order condition (4) then the latter boils down to the linear-in-means behavioral equation of Manski (1993):

$$e_{ig} = \delta_0 + \boldsymbol{\delta}_1 X_i + \boldsymbol{\delta}_2 \overline{X}_g + \delta_3 \frac{a_g}{n_g} + J\overline{e}_g; \tag{6}$$

here $e_{ig} \equiv y_i - y_i^R$ is concealed taxable income, $\overline{e}_g$ is the corresponding average (in a rational expectations sense) in jurisdiction g, $X_i \equiv (y_i, \mathbf{x}_i)$, and $\overline{X}_g \equiv \left(\overline{y}_g, \overline{\mathbf{x}}_g\right)$. The terms $\overline{y}_g$ and $\overline{\mathbf{x}}_g$ denote the jurisdiction-level averages (in a rational expectations sense) of y_i and $\mathbf{x}_i$, respectively.[9] Equation (6) is a reaction function, and J is the main parameter of interest because it

[7] The second-order condition is satisfied as long as $\partial p_{ig}/\partial y_i^R \leq 0$—that is, if $\alpha_0\beta_0 > 0$, which we have assumed to hold.

[8] Corner solutions are possible if the LHS of equation (4) is either strictly greater or strictly less than the RHS. In the first case the taxpayer conceals all of his taxable income; in the second he conceals none. Although such corner solutions are observed in our data (they constitute, respectively, 6.6% and 25.8% of the sample), they are not a source of concern for our empirical analysis. The reason is that such analysis will use jurisdictions as units of observation and all jurisdictions are, empirically, at an interior equilibrium.

[9] The derivation is reported in the Supplemental web appendix (Galbiati and Zanella, 2012).

measures externalities across taxpayers. In Manski's (1993) terminology, J captures endogenous social effects because $\overline{e}_g$ is endogenously determined in the model. All of the parameters in (6) have a structural interpretation. In particular, as shown in the Supplemental web appendix (Galbiati and Zanella, 2012), the endogenous interactions parameter has the following expression:

$$J \equiv \frac{\alpha_1}{2\alpha_0}. \tag{7}$$

The interpretation is straightforward. Recall that α_0 is the parameter determining the extent to which taxpayer i reporting less (more) income drives auditing resources toward (away from) i and that α_1 is the parameter determining the extent to which *other* taxpayers reporting less (more) income drive auditing resources away from (toward) i. Equation (7) expresses the net effect of these two forces on equilibrium behavior. The more sensitive the probability of an audit is to how much other taxpayers in the jurisdiction are reporting, the more a taxpayer reacts to others' behavior. This is the positive dependence of J on α_1. However, such a reaction influences the probability of an audit in the opposite direction, and this affects the expected tax bill in two ways: by altering the expected tax payment on the marginal unit of reported or concealed income and by altering the expected tax payment on all the inframarginal units. The model is linear, so the marginal and the average effect are the same. This explains the negative dependence of J on α_0 multiplied by 2.

The structural interpretation of J, as expressed by (7) is of course model specific, since the only social effect we are modeling is the congestion externality. We have good reasons for this choice, as detailed in Section 2.2. If we allowed for additional social effects, such as an endogenous social norm, then J would also reflect preference parameters. This case is briefly illustrated in the Supplemental web appendix (Galbiati and Zanella, 2012): when the group that defines the social norm coincides with the tax jurisdiction, then one can derive the analogue of (7) for the extended model in closed form and conclude that the parameter we identify is an upper bound of the contribution of the congestion externality. In all the other cases, the interpretation is more complicated.

Notice that the linear-in-means model, equation (6), requires $J < 1$. This is a stability requirement: if $J \geq 1$ then aggregate concealed income would diverge following even a tiny shock. The requirement imposes a restriction on our structural parameters—namely, that $\alpha_1 < 2\alpha_0$. In other words, the congestion externality generated by others' behavior on the probability of an audit, α_1, needs to be sufficiently weak relative to the own effect, α_0. Observe also that $J = 0$ if and only if $\alpha_1 = 0$. That is, if there is no congestion externality then there are no multiplier effects in this model. In that case all social effects, both endogenous (J) and exogenous ($\boldsymbol{\delta}_2$), are absent; in the Supplemental web appendix (Galbiati and Zanella, 2012) we show that if $J = 0$ then $\boldsymbol{\delta}_2 = 0$ as well.

The last step in solving the model is to compute explicitly average concealed income in jurisdiction g. Averaging e_{ig} within a jurisdiction and solving for $\overline{e}_g$ yields

$$\overline{e}_g = \gamma\delta_0 + \gamma(\boldsymbol{\delta}_1 + \boldsymbol{\delta}_2)\overline{X}_g + \gamma\delta_3\frac{a_g}{n_g}, \tag{8}$$

where $\gamma \equiv (1 - J)^{-1}$. The model has a unique Nash equilibrium that is characterized by an individual level of concealed income, a reduced form obtained by substituting (8) into (6), as follows:

$$e_{ig} = \gamma\delta_0 + \boldsymbol{\delta}_1 X_i + ((\gamma - 1)\,\boldsymbol{\delta}_1 + \gamma\boldsymbol{\delta}_2)\overline{X}_g + \gamma\delta_3\frac{a_g}{n_g}. \tag{9}$$

Parameter γ is the social multiplier—that is the ratio between the average cumulative response and the initial individual response following an exogenous shock. If $J > 0$ then $\gamma > 1$ and there are multiplier effects associated with changes in group characteristics or enforcement resources.

2.2 Discussion

There are several important aspects of the model that, because they bear on the interpretation of our empirical findings, merit the more sustained attention provided in this section.

The most important such aspect is that, although we emphasize congestion externalities, other social forces may be at work. First, tax cheating is an activity that requires the development of particular skills: people may learn via social interactions how to conceal their income, or they may find fraudulent accountants more easily in places where tax evasion is widespread (Cowell, 1990, chapter 6). Second, cheating on taxes may violate social norms whose strength decreases with the extent of tax evasion itself (Myles and Naylor, 1996). Third, business informality can be transmitted across the production chain if value-added taxes apply (de Paula and Scheinkman, 2010). Because we do not model these additional mechanisms in the main analysis, one should bear in mind that the structural interpretation of our estimates is model specific (as is often the case in structural estimation). In the Supplemental web appendix (Galbiati and Zanella, 2012) we do sketch an extended version of the model that illustrates how the structural interpretation of the multiplier changes when we allow for preference interactions, but our main framework—with congestion externalities only—is an important benchmark. The reason is twofold. First, although our data allows us to identify precisely the reference groups associated with the congestion externality, it would be impossible (without strong assumptions) to identify the reference groups associated with other types of social effects. Since information about reference groups is fundamental a priori knowledge in empirical studies of social interactions and since wrong assumptions may lead to inconsistency (Conley and Topa, 2003), we prefer to consider only the social effects for which such information is not the outcome of an arbitrary choice. Second, the results of previous research, when combined with ours, suggest that local congestion externalities may actually be the fundamental source of social effects in tax compliance. Fortin, Lacroix, and Villeval (2007) identify endogenous social interactions associated with tax evasion by using laboratory experimental data. They find no significant social effects. A key feature of their experimental design is that audit probabilities are constant at each round and are independent of amounts reported (by oneself and others) across rounds. This feature is important because it ensures fully random and exogenous audit probabilities. At the same time, however, this feature completely shuts off the enforcement congestion channel—that is, the only channel present in our setting. Therefore, the negative finding of Fortin, Lacroix, and Villeval is not at odds with our finding of a large effect; in fact, it helps us to interpret our results. If we assume their experiment is externally valid, then their results and ours taken together—no significant endogenous social effects in a setting with all possible social effects *except* congestion externalities, on the one hand; and significant endogenous social effects in a setting with congestion externalities *only*, on the other hand—imply that the social multiplier we identify reflects primarily the congestion channel.

The second notable aspect of our model is that, contrary to those that introduce nonlinearities (e.g., Brock and Durlauf, 2001a), our model does not produce multiple equilibria. This indicates that multiple equilibria are not needed for generating excess variance. Excess variance may derive—as it does here in a model with a unique equilibrium—from a large social multiplier that amplifies relatively small differences in fundamentals. Such γ-amplification of differences in group characteristics ($\overline{X}_g$ and a_g/n_g) is clear in equation (9).

Third, the social multiplier in our model is a structural parameter representing the short-run, *impact* multiplier effect following an exogenous shock to X_i that induces taxpayers, on average, to conceal one extra dollar. The effect persists in equilibrium as long as the resources a_g of the local tax authorities remain constant. However, it is plausible that these resources adjust in response to shocks that affect equilibrium tax evasion. The speed and the degree of this adjustment determine the dynamic impact of the social externality. In other words, the γ that we identify is the upper bound of the long-run effect. In practice, the long-run effect may be substantially smaller than the short-run one.[10] There is a second type of long-run adjustment that is relevant to understanding of how long the effect of the impact multiplier lasts: taxpayers can move in response to variations in the auditing budget. We believe this second channel is unlikely to be quantitatively relevant in Italy. Mocetti and Porello (2010) analyze the internal mobility of Italian workers and find an annual mobility rate of about 2% since the beginning of the 1990s. It is reasonable to assume that this is also the mobility rate of workers. Mocetti and Porello estimate that, on average, about 22% of workers who moved between 2004 and 2007 were self-employed. Because the self-employed in Italy accounted for 27% of all workers during this period (according to Eurostat), it follows that self-employed workers in Italy have a lower propensity to move than other workers.[11] A rough estimate of their mobility rate is $(.22/.27) \times 0.02 = 1.6\%$. Even if only the individuals with large gains from tax evasion moved strategically across jurisdictions, this mobility rate is so low that it could hardly be a source of concern even in the long run.

A fourth important aspect of our model is the crucial assumption that auditing resources are given, at least in the short-run. This implies that in the short-run the central government can neither increase aggregate resources nor reallocate existing resources across jurisdictions. Several papers have shown that this is true for police resources. For example, Corman and Mocan (2000, 2005) show that adapting police forces to increased criminal activity in New York City takes more than six months. Buonanno and Mastrobuoni (2011) document that, because of a lengthy bureaucratic process, hiring and deploying new police officers in Italy may take up to three years. It turns out that in Italy this is true also for auditing resources: not only

[10] We are grateful to an anonymous referee for emphasizing this point, which is further illustrated in the Supplemental web appendix (Galbiati and Zanella, 2012).

[11] Our interpretation, based on what we know of the socioeconomic environment in Italy, is that the activities of self-employed workers tend to have deep local roots in that country. For this reason the Italian self-employed face above-average mobility costs.

does hiring new tax officers take a long time, but even relocating existing officers or making them work harder may be difficult. We describe such frictions in more detail in Section 3, where we illustrate the institutional background of our data set.

Finally, our model posits a specific social spillover: by observing or expecting more tax evasion at the local level, a taxpayer infers that the probability of detection will be lower until resources adjust. Our model is consistent with alternative mechanisms that have been studied empirically in the literature. For instance, in an investigation of compliance behavior regarding TV license fees in Austria, Rincke and Traxler (2011) identify significant information spillovers from law enforcement. The authors employ compliance microdata that distinguishes between households that were or were not subject to auditing of fee payments. Using snowfall as an instrument for local inspections, they find that households respond positively to increased enforcement in their vicinity. In a static setting like the one we employ here, the rise in compliance among those who had no exposure to field inspections can be interpreted as the effect of a perceived increase in local enforcement resources arising from the observation of greater numbers of audits among neighbors. This is what happens in our model when there is an increase in local resources, a_g.[12] Similarly, in a laboratory experiment of Alm, Jackson, and McKee (2009) analyzes the effect of information concerning enforcement and the compliance behavior of other taxpayers on individual tax-reporting behavior. The authors find that when tax payers are not informed ex ante about the audit probability, information provided by other taxpayers about their audits affects individual compliance. In particular, an increasing number of audits is associated with more compliance. Because the "tax authority" in this lab experiment has no budget constraints the increased compliance can be interpreted as the effect of increased auditing activity on an individual's subjective probability of an audit. Such responsiveness to a higher perceived probability of being audited is also consistent with the findings of Slemrod, Blumenthal, and Christian (2001).

[12] An alternative, dynamic interpretation of the Rincke and Traxler finding within our framework is that past audits increase future compliance of those who were audited, which increases the future audit probability for those who were not. Kleven *et al.* (2011) find that past audits do, in fact, increase complience in a large field experiment in Denmark.

3 Data set

We gained access to the entire collection of tax audits of self-employed individuals in Italy (small individual businesses, including farmers, and professionals) at the end of the 1980s. These are about 80,000 cases.[13] The tax audits we observe were performed at the end of the 1980s on the personal income tax files (i.e., on self-reported income from all sources) for tax year 1987 by the Italian *Guardia di Finanza*, and the results of the audits are now final after all possible appeals. The *Guardia di Finanza* is a "tax police" dependent on the Ministry of the Economy and Finance, and it is in charge of tax enforcement on behalf of the government. In this paper we exploit a crucial feature of the Italian tax audit system: its organization is based on territorial jurisdiction. The Ministry of the Economy and Finance has local branches that are responsible for tax management and audits within well-defined geographic areas. We label these branches "local tax authorities" and the associated areas "tax jurisdictions". At the time our data were collected there were 412 such jurisdictions in Italy. Each jurisdiction comprises one or more municipalities or portions of a large city, and its boundaries are exogenously defined by law following administrative criteria. A local tax authority has no competence beyond its own boundaries. These characteristics imply that tax jurisdictions are the exact reference groups with respect to the endogenous social effect generated by congestion externalities.

Local tax authorities, like the vast majority of administrative entities in Italy, rely on resources allocated by the central government. Our assumption that tax authorities can rely on a given amount of resources in the short run seems realistic in light of what is known about this allocation process. First, hiring new tax auditors implies a long bureaucratic process; even at the local level, hiring a public employee in Italy requires a formal, national-level contest, and it can take years from the day a local tax authority needs additional human resources in the field to the day such resources are actually deployed. As mentioned in Section 2.2, Buonanno and Mastrobuoni (2011) document that, for police officers in Italy, this delay may be as long as three years. Second, the transfer of existing tax auditors from one jurisdiction to another is also a daunting task. In Italy, a public employee must agree to being transferred, and usually no financial incentive is allowed to facilitate that agreement. A confidential interview with an officer at the national tax agency confirms these facts. In particular, the interview reveals that tax

[13] The issue of the representativeness of this data set is briefly addressed at the end of this section and more in detail in the Supplemental web appendix (Galbiati and Zanella, 2012).

auditors are one of the most unionized segments in the Italian public sector (which is itself highly unionized) and that transferring human resources across local tax jurisdictions is subject to bargaining with unions. This picture is in line with our assumption that resource allocation is a sticky process. In addition to geographic information and the sector of economic activity, for each individual we observe income reported for the purposes of personal income tax as well as the amount found by the audit. We take the latter to be an accurate estimate of actual income, and we define *concealed* income as the difference between actual and reported taxable income.[14]

Tables 1 and 2 contain summary statistics disaggregated at the regional level. Table 1 reports information on local jurisdictions and audits. The first column reports the number of jurisdictions; the second reports the fraction of these we classify as "small". A jurisdiction is classified as "small" if the number of self-employed individuals there is less than the average across all jurisdictions and regions. This distinction between small and large jurisdictions is crucial for our identification strategy, as illustrated in Section 4. The third column of Table 1 reports the number of audits, and the fourth column reports the number of self-employed individuals in that region; the last column gives the ratio of these two numbers—that is, the *audit intensity*. This table shows that about 70% of the 412 local tax jurisdictions in Italy are small. As for audit intensity, 2.6% of self-employed individuals were audited at the national level, with some variation across regions: the intensity ranges from 1.9% in Veneto to 3.7% in Molise. Unfortunately, we do not observe the audit intensity at the jurisdiction level.

[14] About 10% of taxpayers in our data *over* report taxable income. Most over reports are of negligible amount, so we treat them as mistakes—consistently with our model—and replace negative values of concealed income with zeros. Such replacement, however, is uninfluential: a robustness check (not reported but available from the authors upon request) shows that our estimates are robust to including all or part of the observed over reports as negative numbers (i.e., negative concealed income).

Table 1. Summary statistics: jurisdictions and audits

Region	(1) Jurisdictions	(2) Small (%)	(3) Audits	(4) Self-employed	(5) Intensity, %
Aosta Valley	2	50.0	267	6,852	3.9
Piedmont	38	73.7	6,887	237,068	2.9
Lombardy	60	73.3	12,634	520,765	2.4
Friuli-V.G.	10	70.0	1,407	65,208	2.2
Trentino-Südtirol	12	100.0	1,067	50,697	2.1
Veneto	31	80.6	4,840	259,584	1.9
Liguria	10	30.0	3,661	95,602	3.8
Emilia-Romagna	24	58.3	6,016	248,353	2.4
Tuscany	34	79.4	5468	215,758	2.5
Marche	14	71.4	2,122	88,906	2.4
Umbria	10	80.0	1,257	43,581	2.9
Lazio	14	35.7	5,729	273,343	2.1
Abruzzo	13	61.5	2,413	66,495	3.6
Molise	4	75.0	569	15,194	3.7
Campania	27	51.8	6,720	228,824	2.9
Basilicata	11	90.1	1,056	27,335	3.9
Apulia	16	31.3	4,968	195,460	2.5
Calabria	31	90.3	2,848	84,175	3.4
Sicily	40	70.0	7,434	217,394	3.4
Sardinia	11	81.8	1,500	73,717	2.0
Italy	412	70.1	78,863	3,014,311	2.6

Notes: Column (1), number of local tax jurisdictions in a given region; column (2) fraction of local tax jurisdictions classified as small (i.e., whose number of self-employed workers is below the average of all jurisdictions; column (3) number of tax audits of self-employed workers; column (4) number of self-employed workers; column (5) audit intensity—that is ratio of column (3) to column (4).

Table 2 reports sample statistics on income and tax evasion. Nominal quantities are deflated using the Italian Consumer Price Index and are expressed in 2010 euros. The first column lists (by region) the average taxable income resulting from auditors' attestations, while the second column lists the corresponding average income resulting from taxpayers' declarations. The third column gives the difference between the two (i.e., our estimate of concealed income) as a percentage of taxable income, and the fourth column reports an alternative measure of tax evasion commonly used in the public finance literature—namely, the tax gap. The *tax gap* is defined as the amount of tax due that was not actually paid, which we express as a percentage of the tax due. Finally, the fifth column reports the average tax rate, defined as the ratio of tax due to taxable income. It is clear from Table 2 that there is a positive relation between taxable income and the average tax rate (the correlation across regions is 0.77), reflecting the progressivity of the tax system in Italy. The table shows that concealed income amounts to 46% of total taxable income in our sample, with considerable variability across regions. The tax gap, instead, is about 55%. These are large numbers, but they are in line with what is known about tax evasion of self-employed workers in other countries. For the United States, Slemrod (2007, Table 1) reports a tax gap of 57% for nonfarm proprietor income (the corresponding figure for farm net income is 72%) and estimates a tax gap of 52% for the self-employment tax. For Denmark, Kleven *et al.* (2011, Table II) report that concealed income is about 42% of taxable income for individuals who are not subject to to third-party report, like in our sample. This last estimate is based on a large set of randomized audits, and is surprisingly in accordance with what we see in our data.

At the individual level, the distribution of concealed income in our sample exhibits wide dispersion even after conditioning on regions (a reasonably homogeneous unit) or tax jurisdictions. A simple ANOVA decomposition reveals that the variance explained by region membership and jurisdiction membership is a mere 0.5% of total variance.

One possible problem with our data set is that it consists of a collection of tax audits, but taxpayers are not (or should not be) audited at random: one expects tax cheaters to be oversampled. In other words, we face a potential selection problem. We claim that this is not an issue: our sample is not much different from a random sample of the underlying population. In the Supplemental web appendix (Galbiati and Zanella, 2012) we substantiate this claim by providing three consistent (and persuasive, we believe) pieces of evidence. Our interpretation is that this fact reflects the inherent difficulty of detecting tax cheaters in Italy back in the 1980s. At that time the tax authority had to rely on what we now know was inadequate information technology. In other words, even though auditing activities were designed to select individuals who were ex ante more likely to evade taxes, the outcome differed little ex post from a random sample.

Table 2. Summary statistics: taxable income and tax evasion

Region	(1) Taxable income	(2) Reported income	(3) Concealed income (%)	(4) Tax gap (%)	(5) Average tax rate (%)
Aosta Valley	40,735	25,741	36.8	41.9	22.6
Piedmont	41,902	26,145	37.6	44.3	23.1
Lombardy	50,377	34,007	32.5	39.2	24.1
Friuli-V.G.	56,287	24,607	56.3	68.5	24.6
Trentino-Südtirol	40,639	22,367	45.0	53.0	24.1
Veneto	54,053	26,834	50.4	62.3	23.6
Liguria	37,002	20,133	45.6	50.1	23.0
Emilia-Romagna	40,255	25,160	37.5	45.6	23.2
Tuscany	41,287	24,492	40.7	48.7	23.0
Marche	31,884	17,054	46.5	52.5	21.8
Umbria	32,629	21,036	35.5	39.9	21.8
Lazio	54,277	25,985	52.1	61.5	23.1
Abruzzo	28,539	15,799	44.6	51.2	21.1
Molise	40,227	13,244	67.1	75.5	22.7
Campania	32,414	14,423	55.5	64.7	21.7
Basilicata	28,796	12,324	57.2	63.8	21.4
Apulia	44,611	14,083	68.4	78.9	22.4
Calabria	32,406	10,973	66.1	77.8	20.1
Sicily	35,066	15,988	54.4	63.4	21.4
Sardinia	35,858	15,729	56.1	64.2	23.4
Italy	41,990	22,503	46.4	55.2	22.8

Notes: Column (1); taxable income, as attested by the auditor; column (2): taxable income, as reported by the taxpayer; column (3) estimate of concealed income (difference between taxable income attested by the auditor and reported by the taxpayer) as a percentage of taxable income; column (4): tax gap (difference between tax due and tax paid) as a percentage of tax due; column (5): average tax rate (ratio of tax due to taxable income). Nominal quantities are expressed in 2010 euros.

4 Identification

4.1 An illustrative example

Because the identification method we use is relatively new, we begin by presenting a simple example to illustrate how identification works and to present in a nutshell the key ideas exploited more formally later in this section. We remark that positive complementarities among decision makers within a given reference group have two important consequences. First, they generate excess variance of individual behavior with respect to individual and group fundamentals. Second, they drive a wedge between the between-group and the within-group variance of individual behavior. These two facts offer a lever for identification. To see this, consider an individual with preferences given by

$$v(e_{ig}; \overline{e}_g) = b_i e_{ig} - \frac{1}{2} e_{ig}^2 + J e_{ig} \overline{e}_g, \tag{10}$$

where e_{ig} denotes a continuous individual behavior (for individual i in group g), $\overline{e}_g$ average group behavior, and J the strength of the complementarity between the two. Think of b_i as the individual-specific, random marginal private benefit of behavior, and think of the quadratic term as the convex cost of such behavior. For simplicity, consider the case in which all groups contain exactly n individuals. Here the unique Nash equilibrium has the linear-in-means form

$$\begin{aligned} e_{ig} &= b_i + J\overline{e}_g, & (11) \\ \overline{e}_g &= \gamma \overline{b}_g, & (12) \end{aligned}$$

where $\gamma \equiv 1/(1-J)$ is the social multiplier and $\overline{b}_g$ is the average of b_i in the group. It is straightforward to verify the two aforementioned properties. First, the excess variance property is easy to see from the ANOVA decomposition of the total variation in individual behavior:

$$\mathbb{V}(e_{ig}) = \mathbb{V}(\mathbb{E}(e_{ig} \mid g)) + \mathbb{E}(\mathbb{V}(e_{ig} \mid g)), \tag{13}$$

where $\mathbb{E}$ and $\mathbb{V}$ denote the expectation and variance operators, respectively. If we substitute equations (11) and (12) into each other and into (13), the result is

$$\mathbb{V}(e_{ig}) = \gamma^2 \mathbb{V}(\overline{b}_g) + \mathbb{E}(\mathbb{V}(b_i \mid g)). \tag{14}$$

The two terms on the RHS have the usual interpretation of, respectively, between-group (explained by group membership) and within-group (unexplained by group membership) variations of individual behavior. There is excess variance because, by (14),

$$\mathbb{V}(e_{ig}) \mid_{\gamma>1} > \mathbb{V}(e_{ig}) \mid_{\gamma=1} . \tag{15}$$

In short, the cross-group variance of mean behavior is larger in the presence of complementarities ($\gamma > 1$) than in the case of independent decision making ($\gamma = 1$). The difference between the two terms in (15) is not explained by fundamentals (i.e., b_i).

Now we verify the second property by writing the within- and between-group variances of behavior explicitly, from (11) and (12), as

$$\mathbb{V}(e_{ig} \quad | \quad g) = \mathbb{V}(b_i \mid g), \tag{16}$$

$$\mathbb{V}(\overline{e}_g) \quad = \quad \gamma^2 \mathbb{V}(\overline{b}_g) = \gamma^2 \frac{1}{n} \mathbb{V}(b_i \mid g) + \gamma^2 \frac{n-1}{n} \mathbb{C}(b_i, b_j), \tag{17}$$

where where $\mathbb{C}$ denotes the covariance operator. Here the last equality follows from the definition $\overline{b}_g \equiv n^{-1} \sum_{i \in g} b_i$, and i and j are generic individuals.[15] Therefore, substituting (16) into (17), we can write the between-group variance as

$$\mathbb{V}(\overline{e}_g) = \gamma^2 \frac{1}{n} \mathbb{V}(e_{ig} \mid g) + \gamma^2 \frac{n-1}{n} \mathbb{C}(b_i, b_j).$$

That is, in the presence of complementarities ($\gamma > 1$) the between-group variance is amplified with respect to the within-group variance. In this linear model, the wedge is equal to γ^2. For the ideal case in which individuals are randomly assigned to groups (which implies $\mathbb{C}(b_i, b_j) = 0$ and also constant within-group variance), γ^2 is identified by $n\mathbb{V}(\overline{e}_g)/\mathbb{V}(e_{ig} \mid g)$. If there is sorting into groups (which implies $\mathbb{C}(b_i, b_j) \neq 0$ and also varying within-group variance) then we need at least one variable z that affects the between-group variance only through the within-group one. That is, z is excluded from the covariance term,

$$\mathbb{V}(\overline{e}_g \mid z) = \gamma^2 \frac{1}{n} \mathbb{V}(e_{ig} \mid g, z) + \gamma^2 \frac{n-1}{n} \mathbb{C}(b_i, b_j),$$

in which case γ^2 can be identified using z as an instrumental variable. This is the key idea in Graham (2008). The identification method that we will employ below is a generalized version of this simple example.

[15] The Supplemental web appendix (Galbiati and Zanella, 2012) provides details about this derivation.

4.2 From linear-in-means to linear-in-variance

The inferential problem is to estimate parameter J in equation (6), and the implied social multiplier, γ. Identification is a concern for a number of reasons, which are now well understood in the social interactions literature (see Blume et al., 2011). First, observe that equations (6) and (8) form a system of simultaneous equations. Identification of the first equation, and so of the endogenous social effect of interest, requires an exclusion restriction in the form of a variable that affects average but not individual concealed income. This is the essence of the "reflection problem" (Manski, 1993)—that is, the problem of separating the effect of mutual influences (endogenous effect γ) from the effect of common influences (contextual and correlated effects $\boldsymbol{\delta}_1$, $\boldsymbol{\delta}_2$, and δ_3) in the reduced form (9). As illustrated by Brock and Durlauf (2001b), the reflection problem could be solved by finding an individual effect (i.e., a variable in X_i) whose average is not a contextual effect (i.e., a variable in $\overline{X}_g$). Since the contextual controls $\overline{X}_g$ are the averages of the individual controls X_i, no such restriction is available. This reflects the open-ended nature of social interactions models, and ours is no exception. Second, even if such an instrument were available, a number of unobservable individual- and group-level effects (including the determinants of self-selection into reference groups) would be confounding factors. The method developed by Graham (2008) considerably mitigates these fundamental identification problems in linear models of social interactions. In this class of models, as the previous example shows, comparing the conditional variance of individual behavior within groups and the corresponding conditional variance between groups allows one to isolate the portion of cross-group variation that is due to endogenous social effects only. This requires that at least one of the conditioning variables does *not* affect certain components of the covariance matrix of individual behavior, a restriction admitting a convenient economic interpretation.

We now illustrate this method in detail for the problem at hand. Consider two vectors W_{1g} and W_{2g} that contain observable jurisdiction-level information, and rewrite the equilibrium equation (9) in variance-components form while assuming that, besides W_{1g} and W_{2g}, we observe only concealed income and group membership. That is, define $\alpha_g \equiv \boldsymbol{\delta}_2\overline{X}_g + \delta_3 a_g/n_g$ as group-level heterogeneity, $\varepsilon_i \equiv \boldsymbol{\delta}_1 X_i$ as individual-level heterogeneity, and $\varepsilon_g \equiv \boldsymbol{\delta}_1\overline{X}_g$ as the group-level average of the latter. Then equations (9) and (8) become

$$
\begin{aligned}
e_{ig} &= \gamma\alpha_g + \varepsilon_i + (\gamma - 1)\,\varepsilon_g, && (18)\\
e_g &= \gamma\,(\alpha_g + \varepsilon_g)\,. && (19)
\end{aligned}
$$

Equation (18) slightly differs from Graham's behavioral equation, where group-level heterogeneity α_g is not amplified by the social multiplier γ. Without amplification, there is no way that that shocks to contextual variables can trigger a feedback chain between group and individual behavior—a chain that is captured by our framework. The reason is that equation (18) is derived from an economic model in which altering institutional variables leads to changes in behavior that are subsequently propagated via social externalities. If one posits a behavioral equation whereby interactions occur directly via average individual characteristics (rather than indirectly in the reduced form, as in our model), then contextual variables that are not reflected in average individual characteristics cannot generate multiplier effects. Graham (2008, 646, fn. 7) recognizes that, in his framework, "γ *may be a composite function of multiple 'structural' parameters. In Manski (1993) it depends on the strength of what he terms 'exogenous' and 'endogenous' social effects.*". Thanks to explicit modeling of the mechanism that generates externalities, the social multiplier we identify has a sharper interpretation: it is a known function of structural parameters and reflects endogenous social effects only.

Denote by $\sigma_\varepsilon^2\,(W_{1g}, W_{2g})$ the variance (conditional on W_{1g} and W_{2g}) of individual heterogeneity, by $\sigma_{\varepsilon\varepsilon}\,(W_{1g}, W_{2g})$ the corresponding conditional covariance across individuals, by $\sigma_\alpha^2\,(W_{1g}, W_{2g})$ the conditional variance of group-level heterogeneity, and by $\sigma_{\alpha\varepsilon}\,(W_{1g}, W_{2g})$ its conditional covariance with individual heterogeneity. Note that $\sigma_{\varepsilon\varepsilon}(\cdot)$ measures the degree of sorting of taxpayers across jurisdictions: it should be zero if they were randomly located with respect to individual characteristics. Similarly, $\sigma_\alpha^2(\cdot)$ captures the variance in unobserved characteristics of the tax authority (e.g., the efficiency of tax auditors, the resources they can rely on) as well as other institutional, cultural, and market characteristics common to all taxpayers in a given jurisdiction. Finally, $\sigma_{\alpha\varepsilon}(\cdot)$ reflects the extent of matching between such characteristics and taxpayers. This covariance is nonzero when, for instance, resources are allocated to tax authorities on the basis of taxpayer characteristics or taxpayers locate themselves across jurisdictions on the basis of how (in)efficient local tax authorities are.

Suppose $\sigma_{\varepsilon\varepsilon}\,(\cdot)$, $\sigma_\alpha^2\,(\cdot)$, and $\sigma_{\alpha\varepsilon}\,(\cdot)$ are all independent of W_{1g}. Following Graham (2008) closely, we show in the Supplemental web appendix (Galbiati and Zanella, 2012) that, after conditioning on W_{1g} and W_{2g}, the within-group variance of concealed income in jurisdiction g (denoted V_g^w) and the corresponding between-group variance (denoted V_g^b) can be written as follows:

$$V_g^w = \mathbb{E}\left[\frac{\sigma_\varepsilon^2\left(W_{1g}, W_{2g}\right) - \sigma_{\varepsilon\varepsilon}\left(W_{2g}\right)}{N_g}\middle| W_{1g}, W_{2g}\right], \tag{20}$$

$$V_g^b = \gamma^2\left(\sigma_\alpha^2\left(W_{2g}\right) + 2\sigma_{\alpha\varepsilon}\left(W_{2g}\right) + \sigma_{\varepsilon\varepsilon}\left(W_{2g}\right) + V_g^w\right). \tag{21}$$

Two points are worth noting. First, the within-group variance (20) is independent of social interactions and group-level heterogeneity. This is intuitive: for example, if there are differences in individual tax evasion in a given jurisdiction in which tax officials are corrupt or there are social externalities, then such variability cannot be ascribed to corruption or social effects (which are the same for everyone, in our model) but only to differences between individuals and to covariance between individual characteristics generated by the process of sorting. Second, the between-group variance (21) depends on group heterogeneity and is amplified by social effects when these are present (i.e., when $\gamma > 1$). This is also intuitive: for example, part of the variability in tax evasion between two groups—one whose tax officials are corrupt and the other not—must depend on corruption. Yet because the level of tax evasion in a group depends on social interactions, which alter contextual differences, so must the cross-group variation. In other words, the presence of social interactions drives a wedge between the variance of illegal behavior at different levels of aggregation. In a linear-in-means model this wedge is proportional to γ^2, which can be exploited to identify γ.

We assume that the portion of the between-group variance (21) that is independent of the within-group variance can be written as a linear function of W_{2g}; thus,

$$\gamma^2\left(\sigma_\alpha^2\left(W_{2g}\right) + 2\sigma_{\alpha\varepsilon}\left(W_{2g}\right) + \sigma_{\varepsilon\varepsilon}\left(W_{2g}\right)\right) = \theta W_{2g}. \tag{22}$$

Next, we rewrite conditional variances as conditional expectations of the appropriate within- and between-group statistics G_g^w and G_g^b, respectively (see the Appendix):

$$V_g^w \equiv \mathbb{E}\left(G_g^w \mid W_{1g}, W_{2g}\right), \tag{23}$$

$$V_g^b \equiv \mathbb{E}\left(G_g^b \mid W_{1g}, W_{2g}\right); \tag{24}$$

then, after using (22), equation (21) becomes

$$\mathbb{E}\left(G_g^b \mid W_{1g}, W_{2g}\right) = \theta W_{2g} + \gamma^2 \mathbb{E}\left(G_g^w \mid W_{1g}, W_{2g}\right). \tag{25}$$

Equation (25) generates a conditional moment restriction,

$$\mathbb{E}\left[G_g^b - \theta W_{2g} - \gamma^2 G_g^w \mid W_{1g}, W_{2g}\right] = 0, \tag{26}$$

which in turn implies the following unconditional moment restriction:

$$\mathbb{E}\left[\begin{pmatrix} W_{1g} \\ W_{2g} \end{pmatrix}\left(G_g^b - \theta W_{2g} - \gamma^2 G_g^w\right)\right] = 0. \tag{27}$$

This equation provides the basis for estimating γ^2 by GMM, with W_{1g} as an instrument. Assuming this is a valid instrument means restricting the covariance matrix of cross-group tax evasion. Therefore, as illustrated by Durlauf and Tanaka (2008), such a covariance restriction parallels the exclusion restriction needed to solve the reflection problem and thus identify social interactions in a regression framework based on model (6). This identification strategy has the important advantage of being robust to arbitrary individual and group-level unobservables.

4.3 The identifying assumption

A "natural" identifying restriction, as in Graham (2008), is provided by group size. The key observation is that the dispersion of individual heterogeneity is greater in small jurisdictions than in large ones. The reason is that individuals with above-average propensity to cheat on taxes are more easily offset in large than in small groups by individuals with below-average propensity to cheat, and vice versa.[16] The maintained identifying assumption is that sorting ($\sigma_{\varepsilon\varepsilon}$), the allocation of auditing resources (σ_α^2), and

[16] This is a standard, testable rank condition:

$$\mathbb{E}\left[G_g^w \mid W_{1g} = 1, W_{2g}\right] \neq \mathbb{E}\left[G_g^w \mid W_{1g} = 0, W_{2g}\right];$$

here W_{1g} is a dummy variable indicating whether the jurisdiction is small or large.

matching ($\sigma_{\alpha\varepsilon}$) are not affected by the size of the jurisdiction. This means ruling out, for instance, that (i) the best auditors are systematically allocated to large (or small) jurisdictions, and (ii) dishonest taxpayers move to small (or large) jurisdictions to avoid such high-quality auditors. Notice, first, that in our data set there is no relation between the size of a jurisdiction and the urban or rural classification of the area where it is located; in particular, large cities contain both small and large jurisdictions. Second, we test a necessary condition that a valid instrument must satisfy: whether the size of the jurisdiction predicts the amount of tax revenue (relative to taxable income) recovered by the local auditor. That amount is a reasonable measure of the auditor's quality and of auditing resources in general. If auditors and resources are allocated based on jurisdiction size, then we should observe a significant correlation between that size and the recovered tax revenue. To check whether this is the case, we regress the jurisdiction-level tax gap on the *small* dummy (see Section 3.1). The results from this regression are reported in Table 3.

Table 3. Relation between jurisdiction size and auditor's quality

	tax gap
small	−0.001
	(0.018)
Constant	0.555
	(0.015)
Observations	412

Note: The dependent variable, *tax gap*, is the jurisdiction-level difference between total tax due (as found by the audit), and total tax paid, divided by the former; *small* is a dummy set equal to 1 if the size of the jurisdiction is below the national average. Robust standard errors are given in parentheses.

The coefficient on *small* is negligible: auditors in large jurisdictions recover only a statistically insignificant 0.1 additional percentage points of the tax gap relative to their colleagues in large jurisdictions (on a basis of 55 percentage points). This supports our assumption that the quantity and quality of auditing resources are not related to the size of the jurisdiction. It follows from this lack of correlation that taxpayers' sorting is also unrelated to the size of the jurisdiction: not only is size unrelated to the urban/rural characteristics of the area but it does not predict the quality of the auditors either. Finally, if neither allocation of resources (σ_α^2) nor sorting ($\sigma_{\varepsilon\varepsilon}$) is related to the size of the jurisdiction then matching ($\sigma_{\alpha\varepsilon}$) is probably also unrelated to *small*. Of course we do not claim that these arguments validate the instrument, but they do weaken the main argument *against* its validity.

Under our identifying assumption, whether the jurisdiction is small or large (i.e., W_{1g}) affects the between-group variance of concealed income only through the within-group variance; this is illustrated by equations (20) and (21). Feasibility requires an estimate of the conditional mean of concealed income, $\mathbb{E}\left(e_{ig} \mid W_{1g}, W_{2g}\right)$. We use the predicted value from the regression of e_{ig} on a constant, on W_{1g} and on W_{2g}. The logic of the model suggests W_{2g} should include information that may affect sorting, matching, and the allocation of auditing resources. A regression of concealed income on the information available in our data set (group size, sector of economic activity, and region) shows that such information is not useful in predicting tax evasion. As a consequence, these variables are not good candidates for inclusion in W_{2g}. We include instead a dummy for whether a region enjoys special autonomy (*Regione a statuto speciale*). There are five such autonomous regions in Italy: Sardinia, Sicily, Trentino-Südtirol, Aosta Valley, and Friuli-Venezia Giulia. These regions have a certain degree of financial autonomy: they retain most of the personal income tax revenue generated locally, and they can impose taxes of their own. Sicily is an extreme example of such financial autonomy: this region not only retains all of the local personal income tax but also collects it directly. In particular, all autonomous regions participate to the tax auditing activities of the central government.[17] Therefore, autonomous region status (dummy variable *special*) is a good candidate for inclusion in W_{2g}. Another good candidate is a dummy for jurisdictions located in Sicily (dummy variable *sicily*), given the extreme case this region represents.

5 Results

Our results are reported in Table 4. Since the model is exactly identified, we simply estimated it by two-stage least squares. The γ^2 that we estimate ranges between 9.52 and 10.46 across the three specifications. The null hypothesis that this is equal to 1 (i.e., $J = 0$) is always rejected at the 10% confidence level. Because this method identifies the square of the social multiplier, it leaves the sign of J undetermined. We assume (consistently with our model) that $J > 0$.

[17] We are grateful to Chiara Martini for pointing out this privilege of autonomous regions in Italy.

Table 4. Results.

	(1)	(2)	(3)
	G_g^b	G_g^b	G_g^b
G_g^w (coefficient: γ^2)	9.54	10.46	9.52
	(4.70)	(5.00)	(4.66)
special	35.45	—	46.16
	(12.42)	—	(20.01)
sicily	—	20.89	−25.14
	—	(12.71)	(22.18)
p-value ($H_0 : \gamma^2 = 1$)	0.07	0.06	0.07
Model parameters (delta method)			
γ	3.09	3.23	3.09
	(0.76)	(0.77)	(0.76)
J	0.68	0.69	0.68
	(0.08)	(0.07)	(0.08)
First stage			
small	4.36	4.13	4.36
	(2.11)	(1.93)	(2.12)
special	−2.09	—	−2.12
	(1.65)	—	(1.84)
sicily	—	−1.90	0.06
	—	(1.42)	(0.76)
F-stat. (excluded instrument)	4.25	4.58	4.23
Observations	412	412	412

Notes: The dependent variable, G_g^b, is defined in equation (**??**) in the Appendix and is such that its conditional mean is the between-jurisdiction variance of concealed income; γ^2 is the square of the social multiplier; *special* is a dummy for whether the jurisdiction belongs to a region that enjoys special autonomy; *sicily* is a dummy for wether the jurisdiction is located in Sicily; *small* is a dummy for whether the size of the jurisdiction is below the national average. Robust standard errors are given in parentheses. The first-stage F-statistic for the excluded instrument (*small*) is robust to heteroskedasticity.

Under this assumption, our estimates imply a value for γ ranging between 3.1 and 3.2 and a value for J ranging between 0.68 and 0.69. We recovered the standard errors of these parameters using the delta method. Our estimates point to a strong amplifying role of the within-jurisdiction enforcement congestion: an exogenous shock altering concealed income independently across individuals produces an equilibrium variation that is up to 3 times the initial response. First-stage results indicate, as expected, that in small jurisdictions the variance of tax evasion is significantly larger than in large jurisdictions; this finding reflects the greater dispersion of individual heterogeneity in small groups.

Although these results cannot be easily compared with other findings in the literature on social interactions, it is worth noting that social multipliers of similar magnitudes have been found in different contexts. Most relevant for our work is that Glaeser, Sacerdote, and Scheinkman (1996) find a social multiplier for crime of 2.8 in the United States. With regard to criminal activity in Italy, Drago and Galbiati (2010) find a social multiplier greater than 2 for the recidivism of former Italian inmates.

6 Conclusion

In this paper we suggested that both the enforcement mechanism and the allocation of resources are important causes of social interdependencies in tax evasion. In particular, endogenous social effects can be generated in a simple model of law enforcement from congestion in auditing resources. Even though social externalities are a plausible explanation for the high variation in tax compliance, empirical research on tax evasion has largely ignored this possible determinant of individual behavior. This state of affairs is due mainly to the extreme difficulty of identifying social effects. We have employed a relatively new identification method (Graham, 2008) that exploits the information contained in conditional variances at different levels of aggregation. This method is particularly useful when working with scant administrative data sets. Using data from Italy and assuming that the size of local tax jurisdictions affects the cross-group variance of tax evasion only through the within-group variance we have identified a tax evasion social multiplier of about 3. We regard this empirical exercise as a contribution to both the tax evasion and the social interactions literatures.

Our work offers important suggestions for tax enforcement policy. An obvious suggestion is that a government can reduce tax evasion at a cost that is much less that the cost of directly inducing each taxpayer to abide by tax regulations. Symmetrically, loosening tax enforcement would reduce tax revenues more when externalities are present. This remark points to another important implication of our findings: tax evasion can be reduced significantly by first removing social externalities among potential tax cheaters. In practice, it is important that fiscal authorities internalize congestion externalities to prevent the perceived probability of punishment from decreasing when more tax evasion is observed at the local level. In particular, our analysis suggests that a more flexible resource allocation mechanism would be an effective tool for controlling tax evasion. The enforcement congestion externality arises because the allocation of resources is a sticky process in the short run. In an ideal system, where resources can be instantaneously allocated, that externality would not be present. This picture is consistent with suggestions from previous studies (e.g., Bordignon and Zanardi, 1997) emphasizing that although the total amount of resources devoted to tax administration in Italy is in line with other countries, the local allocation and use of such resources is inefficient for two reasons: (i) in each region there is a roughly constant, highly persistent number of officers per capita and hardly any mobility across regions; and (ii) a large number of officers are employed in residual administrative tasks while only a small number of them are employed in auditing and enforcement activities. A more flexible allocation process of existing resources would lead to reduced levels of tax evasion at a relatively small cost.

Aknowledgements

We are grateful to Sam Bowles, Bryan Graham, Joel Sobel, Christian Traxler, seminar participants at UCSD, NYU, Amsterdam, Groningen, Siena, Tilburg, and two anonymous referees for useful suggestions. We are particularly indebted to Caroline Hoxby for the insightful discussion of an earlier version of this paper at the 2009 Econometric Society Winter Meeting in San Francisco. The usual disclaimer applies.

References

[1] Allingham, M., and A. Sandmo (1972). "Income Tax Evasion: A Theoretical Analysis." *Journal of Public Economics*, 1(3-4), 323-338.

[2] Alm, J., B. Jackson, and M. McKee (2009). "Getting the word out: Enforcement information dissemination and compliance behavior." *Journal of Public Economics*, 93(3-4), 392-402.

[3] Andreoni, J., B. Erard, and J. Feinstein (1998). "Tax Compliance." *Journal of Economic Literature*, 36(2), 818-860.

[4] Bassetto, M., and C. Phelan (2008). "Tax Riots." *Review of Economic Studies*, 75(3), 649-669.

[5] Blume, L., W. Brock, S. Durlauf, and Y. Ioannides (2011). "Identification of Social Interactions." In J. Benhabib, A. Bisin, and M. Jackson (eds.), *Handbook of Social Economics*, vol 1B, North-Holland, 853-964.

[6] Bordignon, M., and A. Zanardi (1997). "Tax Evasion in Italy." *Giornale degli Economisti*, 56(3-4), 169-210.

[7] Brock, W. and S. Durlauf (2001a), "Discrete Choice with Social Interaction", *Review of Economic Studies*, 68, 235-260.

[8] Brock, W. and S. Durlauf (2001b), "Interactions-Based Models." In J. Heckman and E. Leamer (eds.), *Handbook of Econometrics*, vol 5,, Elsevier, 3297-3380.

[9] Buonanno, P., and G. Mastrobuoni (2011). "Police and Crime: Evidence from Dictated Delays in Centralized Police Hiring." *Working paper*, University of Bergamo and Collegio Carlo Alberto.

[10] Conley, T., and G. Topa (2003). "Identification of local interaction models with imperfect location data." *Journal of Applied Econometrics,* 18(5), 605–618.

[11] Corman, H. and N. Mocan (2000). "A Time-Series Analysis of Crime, Deterrence and Drug Abuse in New York City." *American Economic Review,* 90(3), 584-604.

[12] Corman, H. and N. Mocan (2005). "Carrots, Sticks and Broken Windows." *Journal of Law and Economics*, 48(1), 235-266

[13] Cowell, F. (1990). *Cheating the Government. The Economics of Evasion.* MIT Press.

[14] de Paula, A., and J. Scheinkman (2010). "Value-Added Taxes, Chain Effects, and Informality" *American Economic Journal: Macroeconomics*, 2, 195-221.

[15] Drago, F., and R. Galbiati (2010). "Indirect Effects of a Policy Altering Criminal Behaviour: Evidence from the Italian Prison Experiment". IZA Dp. 5414.

[16] Durlauf, S., and H. Tanaka (2008). "Regression Versus Variance Tests for Social Interactions." *Economic Inquiry*, 46(1), 25-28.

[17] Ehrlich, I. (1973). "Participation in Illegitimate Activities: A Theoretical and Empirical Investigation." *Journal of Political Economy*, 81(3), 521-565.

[18] Ferrer, R. (2010). "Breaking the law when others do: A model of law enforcement with neighborhood externalities." *European Economic Review*, 54, 163-180.

[19] Fortin, B., G. Lacroix, and M. Villeval (2007). "Tax Evasion and Social Interactions." *Journal of Public Economics,* 91(11-12), 2089-2112.

[20] Frey, B., and L. Feld (2002). "Deterrence and Morale in Taxation: an Empirical Analysis." *CESifo Working Paper* No. 760.

[21] Galbiati, R., and G. Zanella (2012). "Supplement to 'The Tax Evasion Social Multiplier: Evidence from Italy'." Online appendix available at the authors' websites.

[22] Glaeser, E., B. Sacerdote, and J. Scheinkman (1996). "Crime and Social Interactions", *Quarterly Journal of Economics*, 111, 2, 507-548.

[23] Gordon, J. (1989). "Individual Morality and Reputation Costs as Deterrents to Tax Evasion." *European Economic Review*, 33(4), 797-805.

[24] Graham, B. (2008). "Identifying Social Interactions through Conditional Variance Restrictions." *Econometrica*, 76(3), 643-660.

[25] Judt, T. (2005). *Postwar: A History of Europe Since 1945*. Pimlico.

[26] Kleven, H. J., M. Knudsen, C. Thustrup, S. Pedersen, and E. Saez (2011). "Unwilling or Unable to Cheat? Evidence From a Tax Audit Experiment in Denmark." *Econometrica*, 79(3), 651–692.

[27] Manski, C. (1993). "Identifications of Endogenous Social Effects: The Reflection Problem." *Review of Economic Studies*, 60, 531-542.

[28] Manski, C. (2000). "Economic Analysis of Social Interactions." *Journal of Economic Perspectives*, 14(3), 115-136.

[29] Mocetti, S., and C. Porello (2010). "La mobilità del lavoro in Italia: nuove evidenze sulle dinamiche migratorie." *Bank of Italy Occasional Papers.*

[30] Myles, J., and R. Naylor (1996). "A Model of Tax Evasion with Group Conformity and Social Customs." *European Journal of Political Economy*, 12, 49-66.

[31] Pisani, S., and C. Polito. (2006). "Analisi dell'Evasione Fondata sui Dati IRAP. Anni 1998-2002." Agenzia delle Entrate–Ministero Economia e Finanza.

[32] Rincke, J., and C. Traxler (2011). "Enforcement Spillovers." *Review of Economics and Statistics*, 93(4), 1224-1234.

[33] Sah, R. (1991). "Social Osmosis and Patterns of Crime." *Journal of Political Economy*, 94, 1272-1295.

[34] Sanchez, I. and J. Sobel (1993). "Hierarchical Design and Enforcement of Income Tax Policies." *Journal of Public Economics*, 50(3), 345-369.

[35] Sandmo, A. (2005). "The Theory of Tax Evasion: A Retrospecive View." *National Tax Journal*, 58(4), 643-663.

[36] Schlicht, E. (1981). "Reference Group Behaviour and Economic Incentives: A Remark." *Zeitschrift für die gesamte Staatswissenschaft*, 137(1), 125-127.

[37] Scotchmer, S. (1987). "Audit Classes and Tax Enforcement Policy." *American Economic Review*, 77(2), 229-233.

[38] Slemrod, J. (2007). "Cheating Ourselves: The Economics of Tax Evasion." *Journal of Economic Perspectives*, 21(1), 25-48.

[39] Slemrod, J., M. Blumenthal, and C. Christian (2001). "Taxpayer response to an increased probability of audit: evidence from a controlled experiment in Minnesota." *Journal of Public Economics*, 79(3), 455-483.

[40] Yitzhaki, S. (1974). "A Note on 'Income Tax Evasion: A Theoretical Analysis'." *Journal of Public Economics*, 3(2), 201–202.