# Tax Professionals and Tax Evasion*

Marco Battaglini[†] Luigi Guiso[‡] Chiara Lacava[§] Eleonora Patacchini[¶]

In the traditional literature on tax compliance, tax advisors provide technical expertise aimed at clarifying and interpreting legislation. We show evidence that some tax advisors play a more expansive role, providing services to taxpayers who intend to evade taxation. We study this role with an exclusive data set covering the entire population of sole proprietorships in Italy, the respective audit files and tax advisors. Exploiting quasi-random variation in audit policy, we document that tax advisors act as information hubs, gathering privileged information on the auditing policy from their activity and incorporating it into the tax return strategy of their customers. Heterogeneity in the accountants' willingness to serve this role establishes a market for intermediated tax evasion in which taxpayers sort themselves on the basis of the tax advisors' tolerance for it. These findings have important implications for the design and evaluation of audit policies.

*Keywords*: tax enforcement, tax compliance, tax evasion, indirect treatment effects

*JEL classification*: K34, H26

* A previous version of this paper based on a subset of the data circulated with the title: Tax Professionals: Tax Evasion Facilitators or Information Hubs? We are very grateful to the Italian Revenue Agency for granting us access to the data. We are solely responsible for the ideas expressed in the paper. We thank Marius Brülhart, Stephen Coate, Bernard Fortin, Roberto Galbiati, Philipp Kircher, Doug Miller, Matteo Paradisi, Alex Rees-Jones, Giulio Zanella, and participants to several seminars for valuable discussions.

† Cornell University and EIEF E-Mail: battaglini@cornell.edu

‡ EIEF E-Mail: luigi.guiso55@gmail.com

§ Goethe University Frankfurt E-Mail: lacava@econ.uni-frankfurt.de

¶ Cornell University E-Mail: patacchini@cornell.edu

# 1 Introduction

The traditional literature on tax evasion focuses on the direct relationship between the tax authority and taxpayers: individuals are assumed to be independent utility maximizers who trade off costs and benefits of violating tax laws. It is, however, increasingly the case that the relationship is more complex, as it is mediated by tax professionals. This evolution is caused by the increasing complexity of the tax code within each country, and new opportunities for avoidance and evasion between countries thanks to globalization — at least for corporations and the upper tail of the income and wealth distribution. In this context, understanding the role of tax professionals is key to minimizing the cost of compliance and making auditing more efficient. What do tax professionals do *beyond* helping their clients understand and apply the laws?

A recent sociological literature suggests that tax professionals play a key role in the formation of their client's expectations regarding enforcement probabilities and shape tax norms and ethical standards (Smith and Kinsey, 1987, Braithwaite, 2005, Raskolnikov, 2007, Harrington, 2016). The literature in economics, however, has focused on the role of tax professionals as providers of technical expertise —i.e. as tax advisors. Studies have been done on what determines the choice of hiring a practitioner and of its type, on the value of practitioners in reducing the time and anxiety cost associated with tax return preparation, and in assessing the benefit of the practitioners' expertise in interpreting the norms.[1] The study of whether and how tax practitioners affect tax compliance has been constrained by data availability, since it is rarely the case that detailed information is available on taxpayers, their tax accountant and audits.

In this paper, we contribute to this literature by studying the role of tax accountants (TA for brevity) in tax compliance with a new, exclusive data set on the entire population of sole proprietorship in Italy, the respective audit files and TAs. To guide our inquiry, we propose a simple theoretical framework which suggests that, besides providing technical expertise on how to interpret tax laws, TAs act as *informational hubs*: by serving their clients, they naturally collect detailed information on the auditing strategy, one of the most guarded secret of any tax authority. This information does not only include the frequency of the audits, but detailed knowledge of what the tax authority audits and how the audits are conducted. The TA incorporates the information acquired from each individual client onto the services provided to all clients, thus transforming the nature of the interaction between the tax authority and each individual taxpayer from a direct and independent relationship, to a mediated, social relationship in which the behaviors of all clients of the same TA are

[1] We discuss in greater detail this literature in Section 1 below.

connected. The analysis also suggests that this valuable information and the associated services create a *market for intermediated tax evasion,* as the taxpayers sort themselves out in order to match their willingness to evade with the willingness and ability of the TA to enable their evasion.

Our data are unique to assess this theory. They are a combination of individual level information from two separate administrative records from the Italian tax authorities: the return files for the universe of sole proprietorship taxpayers (approximately 4.7 million) from 2007 to 2013, and the respective audit files. The data provide detailed information about the taxpayer's reported income, demographic characteristics, business characteristics, and audit history (including the outcome of any audit), and detailed information on the TA employed by the taxpayer. Around 97% of taxpayers in our sample rely on the services of a TA, who is identified by a unique code by law. This allows us to match taxpayers with accountants and follow the history of the taxpayer-accountant relationship over time. Two additional exclusive features of our data set play a key role in our analysis. First, we have information on the specific authority initiating the audit. Exploiting this, we are able to identify a set of audits that are random conditioning on observables. This allows us to address the natural concern when studying tax enforcement using administrative data that audits are not randomly assigned across taxpayers.[2] Second, we are provided unique private information on the TAs themselves: their personal tax returns, turnover, profitability and, critically, the outcome of audits on them.

While the behavior of tax advisors as tax evasion intermediaries is clearly affected by the institutional set up of the economy where they operate, their role as information hubs and the tax payers incentive to match with accountants of their type are general features of any economy. Thus, even if our investigation leverages on data from a specific country the qualitative findings should apply broadly.

We start our analysis documenting a positive and robust correlation between taxpayers' evasion and that of their TA, as measured by the average evasion of her/his other clients over the entire period under analysis. We show that a taxpayer's evasion is also correlated with their own TA's tendency to evade, and that this correlation does not depend on the TA's other characteristics, such as experience, quality of services, or specialization in terms of clients' business sector.

[2]In Italy tax audits can be initiated by two tax authorities, the *Italian Revenue Agency* (IRA), an administrative body in charge of tax collection and enforcement, and the *Guardia di Finanza* (GdF), a police force with a wide range of responsibilities including tax enforcement. While the latter naturally relies on "soft information" that we do not observe in the selection of audits, the former is in charge of a program of automated and desk assessments that are based on data in the *Anagrafe Tributaria* (the official Tax Registry), for which we have been granted exclusive access. As we show in Section 2.3, this difference allows to establish that the audits originated by the IRA are random conditioning on observables in our data set.

We then study the role of TAs in tax compliance by looking at whether TAs play an active role in collecting and distributing information on tax audits. We look at the difference between the income reported by a taxpayer $i$ if other clients of his/her accountant $j$ have been audited in the previous year and that of the same taxpayer if the peers happen to receive no audit. We find that an audit at $t-1$ to at least one other customer of accountant $j$ induces an increase of 1.3% of $i$'s reported income at $t$. This indirect effect is additional to a 7.5% direct effect of own audit — that is to the increase in the income reported at $t$ induced by an audit of $i$ at $t-1$. Critically, we show that the indirect effect passes through the TA, and it is not due to direct interactions between the clients. First, we show that the spillover is still present and similar in magnitude for clients of the same TA who are different and thus less likely to interact directly: for example, we find virtually identical spillovers for clients of the same TA who are in different provinces and/or operate in different business sectors. Second, we show that no spillover effect is present when we look at audits on taxpayers with similar observed characteristics but different accountants. Third, we rule out that the spillover is due to the presence of social links uncorrelated to observable characteristics (for example, past experiences, friendship, kinship). Because our data cover the universe of taxpayers over time, for taxpayers that switch TA over the sample period, we observe reported income and audits before and after they join a TA. If a taxpayer is socially connected to other clients of a TA (through links of friendships or family), then those links should be at work before s/he hires the TA, and after s/he leaves the TA too. We look at the effect of an audit at $t-1$ to a taxpayer who changes TA at $t$ on the reported income of the customers of the TA at $t$. Similarly, we look at the effect of an audit at $t+1$ to a taxpayer who changes TA at $t$ on the reported income of the customers of the TA at $t$. For both cases, we show that there is no effect of audits on other clients of the same TA before a taxpayer joins a TA, or after a taxpayer leaves a TA.

We next reveal the presence of a market for *intermediated tax evasion* by showing evidence of self-selection of taxpayers, sorting themselves into accountants of similar tolerance for tax evasion. To this goal, we follow taxpayers as they, voluntarily or involuntarily, switch TA. If taxpayers self-select into a TA of their type, we expect the type of their accountant before the switch (as measured by the TA own evasion and the observed historical tendency of its clients to evade) to be correlated with the type of their accountant after the switch. We show this is indeed true both in the entire sample of taxpayers who switch TA, and in the sub-sample of taxpayers who are forced to switch because their TA exits the market (due to, e.g. retirement or death). We show that sorting on tax evasion is not a spurious reflection of sorting on other relevant characteristics of the taxpayer that are potentially correlated with tax evasion, particularly the business sectors. In fact, we show that TAs are not specialized in serving clients in a specific business sector, and we control for taxpayers' sector and location using a very fine partition with over 1,200 sectors and over 8,000 municipalities. We also show that taxpayers sort along other key characteristics of the TA that are functional or ancillary to tax evasion. Our data allow us to construct an indicator of the TA's tax avoidance (measured using the difference between gross income and taxable income of clients), and an indicator of how aggressive the TA is in defending the clients in case of a successful audit (measured by the rate of appeal in case of detected evasion). We document that after a taxpayer switches accountant, s/he tends to select a TA with similar indicators to the old TA, both for voluntary and involuntary switches.

The results described above have important policy implications, both in terms of the evaluation of audit deterrence effects, and in terms of lessons on how to design an audit system. The first lesson is that accountant-induced peer effects tend to magnify the effect of audits along two dimensions: an audit on $i$ has an effect on *all* peer-customers of $i$'s accountant; the peer effects, moreover, tend to persist over time. The effect on $i's$ currently reported income of a three-year old audit on other customers of the same accountant is not only positive but even larger than the one-year old effect. These results suggest that an audit has a powerful deterrent effect on other taxpayers in the relevant "social circle" of the audited taxpayer. Because of this, they also suggest that audits for subsequent years on clients of the same TA are less likely to find evasion, since the taxpayers responded by evading less. The second lesson is that when evasion is found for one client of a TA, other clients are likely to have evaded as well: the tax authority should incorporate this information in the auditing strategy. This point is not in contradiction with the previous, since the tax authority can audit income produced in years preceding the audit of a peer, thus exploiting that the sorting occurred pre-audit and avoiding a possible change in behavior among other clients post-audit.

The data show that none of these lessons are currently exploited by the Italian tax authority. Finding that a taxpayer (or a TA) is an evader does not affect the rate of auditing of other clients of the same TA, which our results suggest should be increased; nor does it affect the auditing strategy of other clients of the same TA for following years, which our results suggest can be reduced because of the deterrence effect.

We proceed as follows. After relating our paper to the literature below, in Section 2, we provide an overview of our data and institutional background. In Section 3 we portray in detail TAs and their clientele. In Section 4, we show evidence on the correlation between a taxpayer's income evasion and that of the other clients of her/his TA, documenting its robustness to a vast set of model specifications and estimation strategies. Guided by a simple model of the interaction between taxpayers and TAs (detailed in the Appendix), in Section 5 we study empirically the mechanisms through which the correlation can be generated, and show the existence of a market for intermediated evasion. Section 6 discusses the policy implications of our findings, and Section 7 concludes.

**Related literature.** Our work combines two literatures that have remained mostly separated up until now. First, the relatively small literature on tax practitioners, and second, the literature on social spillovers in tax compliance. Research on tax practitioners has traditionally focused on their role as providers of expert advice but has ignored the potential social spillovers between clients of the same practitioner, implicitly assuming that the TA does not change the direct relationship between tax authority and a taxpayer.[3]

Recent contributions have provided persuasive evidence that practitioners may improve the ability of taxpayers to understand tax rules.[4] The focus of these papers, however, is on how heterogeneous knowledge of the tax law affects labor supply and tax reports. This literature has not attempted to distinguish whether a TA's expertise leads to a legitimate and more efficient use of information, or on more knowledgeable tax avoidance strategies.[5] Contrary to these works, we have detailed knowledge of tax audits, their outcomes and, importantly, on the TAs. This information allows us to study how TAs help taxpayers respond to individual audits on their clients — as opposed to just helping them navigate tax law more efficiently — and how clients sort themselves in terms of the tax evasion of other clients. Because we observe the tax evasion of the TAs, we can study how the information hub activity and the sorting depends on the tax evasion propensity of the advisor in a clean and direct way.

The literature on social spillovers has focused on showing network externalities in compliance behavior, relying on lab or field data where the network is mainly identified by spatial proximity.[6] Some evidence that the affiliation to a common tax practitioner may also be

[3]The focus instead has fallen on the determinants of hiring a practitioner or not (Erard, 1993), the usefulness of tax practitioners (Slemrod, 1989), the effect on the level and type of compliance (Klepper et al., 1991, Erard, 1993), and the role played by practitioners in reducing uncertainty and costs of compliance (Scotchmer, 1989, Beck and Jung, 1989, Reinganum and Wilde, 1991). See Andreoni et al. (1998) for a survey of this literature.

[4]See in particular Chetty et al. (2013), Chetty and Saez (2013) and Zwick (2021).

[5]Indeed, Chetty and Saez (2013) and Zwick (2021) show that taxpayers tend to underutilize even legitimate tax rules, and that advisors provide legitimate advice to reduce this inefficiency. Bunching — the phenomenon documented by Saez (2010) and Chetty and Saez (2013) where taxpayers tend to report income at levels corresponding to kinks of the tax rate — is not necessarily a sign of cheating, but it could reflect a natural and efficient response of endogenous labor supply to the non-linearity of the budget constraint generated by changes in marginal tax rates (Saez, 2010, Garin et al., 2021).

[6]Early work tested the hypothesis of social spillovers using laboratory experiments (see Fortin et al., 2007, Alm et al., 2009, and Alm et al., 2017). More recent research has used field data and experiments to investigate the role of spatial proximity. See for instance Rincke and Traxler (2011), Galbiati and Zanella (2012), Del Carpio (2014) and Perez-Truglia and Troiano (2018).

a source of network effects can be found in Boning et al. (2020). This paper studies the comparative effect of either an informational visit by an internal revenue officer or of an informational letter from the IRS on firms suspected of noncompliance with the requirement to remit withheld income and payroll taxes, which are due every quarter. It documents an effect on firm $i$'s remittances of direct visits to other firms served by the same tax practitioner as $i$'s. Their data however do not allow to tell whether the spillover occurs through the TA or rather through the social network that may connect the customers of a given TA: with a policy experiment it is possible to randomize audits, but it is still impossible to randomize the affiliation with TAs, which remains endogenous. Contrary to this paper, we can exploit the panel structure of our dataset to prove the role played by TAs and rule out that it captures social interaction. We indeed observe a treatment (an audit) before and after a taxpayer joins a TA, so we can observe if the audit has an effect on other clients before and after the targeted taxpayer becomes a client. In addition, the size and longitudinal structure of our data allows us to provide evidence of sorting of taxpayers seeking tax evasion services — a key feature to pin down the role of TAs as tax evasion intermediaries, that has never been studied in the literature before. Differently from Boning et al. (2020), we observe audits and their outcomes for the entire population, including TAs. As a result, we have the unique ability to study the existence and functioning of a market for intermediated evasion and the pivotal role of the TA in it.[7]

# 2 Institutional background and data

## 2.1 Institutional background

In Italy the administration of tax revenues is decentralized to two distinct agencies: the Italian Revenue Agency (*Agenzia delle Entrate*, IRA), an administrative body in charge of tax collection and tax enforcement; and the Financial Police (*Guardia di Finanza*, GdF in brief), a military police force responsible for dealing with financial crime, smuggling, customs and borders checks, and patrolling Italy's territorial waters. GdF contributes to tax

[7]An important related but distinct literature studies the role of third-party reported paper trails for enforcement. Taxpayers whose income is reported completely or partially by a third party have lower incentives to evade because their evasion can be easily detected (Kleven et al., 2011, and Pomeranz, 2015). Pomeranz (2015) shows that companies that generate a VAT paper trail respond less to exogenously generated changes in their perceived audit probability. Changes in their perception, moreover, increase VAT payments to their suppliers. Pomeranz (2015) emphasizes production linkages across firms as a vehicle through which tax audits can spillover. We focus on spillovers generated by the common TA even among firms that would otherwise be unrelated. Common to both is the importance of accounting for taxpayers interconnections when assessing tax enforcement.

enforcement collaborating with the IRA.[8] Both agencies can initiate a tax audit but each follows different methodologies that leverage their comparative advantage.

The IRA is almost exclusively in charge of designing the program of automated assessments that are routinely carried out by the agency based on information and data available in the Tax Registry *(Anagrafe Tributaria)*, elaborated via a number of different applications.[9] The Tax Registry is a centralized database that identifies each taxpayer with a unique tax code (*Codice Fiscale*, the analog of the US Social Security Number) and associates it with a rich set of statistical information. Audits are chosen based on these records. More specifically, since the institution of the *Studi di Settore* (Sector Studies) in 1993, the Italian legislature has formalized the idea that IRA tax audits of medium-small businesses and practitioners should be strongly inspired by conformity of these types of taxpayers to indicators of expected economic performance and presumptive income.[10] The latter is defined at a narrow geographic and business sector level by a committee of experts on the basis of statistical studies. These studies gather information on homogeneous groups of taxpayers and provide estimates of (i) the minimum income they should produce and that are thus expected to file, and (ii) the range of several indicators of economic performance (D'Agosto et al., 2017). Based on these studies, for each tax filing the IRA elaborates two indexes of "conformity" to *Studi di Settore* that are used to target individual audits: an indicator variable called "congruent", and one called "coherent." A "congruent" tax filing is one whose reported income is above the estimated minimum income threshold identified in *Studi di Settore*. A "coherent" tax filing is one with no indicator of economic performance outside the estimated ranges computed by *Studi di Settore* for a taxpayer of that type. Non-congruent and non-coherent files have a higher probability to be audited. Both indicators, as well as other information used by the IRA to select audits, are in the data that the IRA has exclusively shared with us and that we describe in greater detail below. This gives us a direct insight into the IRA's auditing policy. The GdF, on the other hand, has a widespread presence in the territory, suitable for carrying out in-depth investigations aimed at uncovering a wide spectrum of illegal activities. This gives the GdF exclusive access to "soft" information which can be exploited to design its audit policy.

In sum, while IRA audit policy relies on "hard data", the GdF audits are guided by insights from information gathered as part of its police activities. Below we exploit this institutional arrangement to identify a set of quasi-random audits, that are random conditional on the statistical information available to the IRA when the audits are selected.

[8] A third agency involved with tax audits is the *Custom Agency*, but this agency has a minor role and is not relevant to taxpayers in our sample.

[9] See OECD, 2016, p. 48.

[10] See law n. 427/1993 that introduced the *Studi di Settore* and law 146/1988 that regulates their use.

## 2.2 Data description

Our study relies on population-level data for all Italian sole proprietorship businesses. We merge information from two different administrative records from the IRA: returns files and audit files. In both cases, records are at the individual level and cover filings of incomes generated in seven fiscal years, from 2007 to 2013, reported between 2008 and 2014, and audited between 2009 and 2015.

The data contain detailed information on all components of taxpayers' tax returns (including reported taxable income, turnover, liabilities and deductions), their demographics (gender, age, marital status), and characteristics of their business (years of activity, number of employees). Importantly, we were granted access to the detailed geographical location where the business operates (8,054 municipalities, with a median of 125 sole proprietor taxpayers per municipality), and to the highest level of disaggregation of the sector of activity (ATECO 5-digit code, that is 1,215 sectors, with a mean number of taxpayers per sector of 2,390).[11] The audit data contain information on whether and when the taxpayer was audited, which filing was audited, and by which agency (IRA or GdF). The data contain the result of the audit, with assessed taxable income and the amount of evasion found (if any) computed as the difference between reported and assessed income.[12] We also observe whether there was an in-person visit of an inspector and if the taxpayer appealed against the audit. Importantly for our purposes, the data report the identifier of the TA who filed the taxpayer's tax statement.[13] Because we have population-level data we can trace all taxpayers that are served by the same TA, observe taxpayers' mobility from one TA to another, as well as TA closures which force taxpayers to match with a new TA.

The taxpayers in our data set are individuals who own a sole proprietorship, where no legal distinction is made between the enterprise and the sole owner. Table 1, panel A shows summary statistics. Overall, our sample contains almost 4.7 million taxpayers and 20.3 million filings. About 27% of the taxpayers are women, 65% are married, and the average age is 47 years. The average enterprise has been in operation for 13 years and, consistent with the small size of these businesses, employs 0.8 workers with relatively limited variability

[11] To give an idea of the level of detail of the sectoral information, for retail sale of clothing we can distinguish between sales of clothing for adults, clothing for children, underwear, articles of fur or leather, and clothing accessories. Or for design activities we can distinguish between fashion design, industrial design, web designers, illustrators, graphic designers, and interior decorators.

[12] Audits' assessed evasion can be negative if people file income in excess of their true taxable income. Audits initiated by the GdF are transmitted to the IRA, which collects all audit data in its archives.

[13] We drop tax returns filed without the advice of a TA (2.5% of the observations). Sole-proprietorships might have additional income as employed workers (this is the case for 28.4% of the observed filings). Our working definition of taxable income is the difference between the total taxable income and the additional income as employed workers. We drop filings reporting a taxable income as employed workers larger than their taxable income as sole-proprietors (4.2% of the observations).

(standard deviation 3.2, 90th percentile: 2 employees). About 37% of our taxpayers provide services in the private sector (e.g. lawyers, hairstylists, coffee shop owners, architects), 4% provide health, education and recreational services (e.g. physicians, dentists), 19% are in constructions and manufacturing (e.g. small construction firms, plumbers, artisans, bakers), 27% in trade and 12% in agriculture. The average reported taxable annual income is 18,640 euros with relevant heterogeneity (standard deviation: 48,694 euros; 90th percentile 39,997 euros), partly reflecting differences across industries. While the taxable income is reported net of deductions, the tax statement also reports the gross income (before deductions). We construct a proxy measure of tax avoidance as the difference between gross annual income and net taxable income, and scale it by gross income.[14]

During the sample period, 289,434 taxpayers (6.2% of the total) were audited at least once. Table 1, panel B describes the audited tax filings. The vast majority of the audits were originated by the IRA (97%) and the rest by the GdF . The share of audited tax filings is around 1.9%. However, there is some variation over time and heterogeneity in auditing probabilities across sectors, even though differences are mild (only taxpayers in agriculture are associated with an audit probability which is significantly lower than average). The tax authority uses the discrepancies between filed income and the conformity indexes (elaborated in the *Studi di Settore* discussed in the previous section) as evasion-risk indicators. Around 35% of the filings are classified as non-congruent and 52% as non-coherent, with some variation over different fiscal years. Tax filings of larger firms, defined either in terms of the number of employees or the value of filed income, are more likely to be audited: the filed taxable income of an audited filing is 1.6 times the average filed income. Usually, tax filings are audited 4 years after their filing and the audit lasts less than 4 months, with a median of 2 months and more than 6 months for around 10% of the audits. Conditional on being audited, the fraction of filings with positive evasion is 66.45%. Conditional on positive evasion, the average income evaded amounts to 32,689 euros, 1.1 times the average income filed. The size of the share evaded and the level of evasion are both quite dispersed (as shown in the two panels of Figure S4 in the Online Appendix); larger evasions, however, account for a substantial fraction of the total (the evasion of the top 10% accounts for 67% of the total evasion). The average share of evasion is 0.33 and, despite the large variation in the filed income, the share of evasion is similar across industries (between 0.30 and 0.38), except for the health, education and recreational services activities (0.18). Around 7.1% of audited taxpayers are found to have evaded all of the taxable income, and on average taxpayers with

[14] Average tax avoidance is around 21%, with significant differences across industries. Businesses in agriculture and retail have higher than average tax avoidance (26% and 23% respectively), while services have below average tax avoidance (19% for private services and 12% for health, education and recreational services).

positive evasion evade half of their taxable income. After receiving an audit, around 19% of the taxpayers appeal the result of the audit assessment, with low variation across industries (the appeal rate ranges from 15% of the cases in construction and manufacturing to 23% in agriculture).

In addition to these data, the Tax Authority provided us with the identifier of the local office in charge of each filing (288 in total). This additional (and unique) information is important for our identification strategy since it allows us to control for differences in audit capacity (e.g., number and experience of inspectors, budget allocated to enforcement activities) across offices.

## 2.3 Quasi-random audit selection

Our data set includes millions of data points on taxpayers, including their demographic characteristics, balance sheet indicators of their business activities, and information on their TAs (discussed in Section 3). Since the IRA decides which tax filing to audit with the same hard information available to us, we obtain random variation in the IRA audit selection if we use a sufficiently large set of conditioning variables. In this section, we show that after conditioning on a selection of the available data, audit decisions are effectively random, thus ruling out that they are driven by additional "soft information" that we do not observe.

To this goal we propose two sets of tests. First, we show that the selection of variables on which we condition our analysis is sufficient to describe the IRA's audit policy. Naturally, the IRA has not provided us with the exact algorithm they use to identify the specific tax filing to be audited, but has shared the output of their main elaborations from the *Studi di Settore* and have guided us in the selection of the other relevant variables. We use a large number of variables including characteristics of the tax filing (the indexes of conformity from the *Studi di Settore* discussed in the previous section, fixed effects for number of years passed since filing); characteristics of the taxpayer and its firm (gender, marital status, years of activity, sector of activity, number of employees); and characteristics of the tax practitioner (including the number of clients and geographical dispersion of the clients). We also include year of filing fixed effects (capturing common time trends in evasion), location fixed effects (picking up systematic differences in the propensity to evade across areas due e.g. to cultural differences), and business sector fixed effects (reflecting e.g. differences in ease of hiding income across sectors). To investigate whether these variables represent audit decisions well, we estimate the probability of being audited with our selection of variables and with variables selected by machine-learning techniques, showing that the two approaches have comparable prediction accuracy. We identify the relevant variables among the ones available in our data (more than

230) using the Least Absolute Shrinkage and Selection Operator (LASSO); and we then form a prediction using the selected variables.[15]

Table 2 reports goodness of fit statistics for LASSO probit predictions obtained using alternative shrinkage estimators. For each model, we report the number of non zero coefficients, the out-of-sample deviance, and the deviance ratio.[16] The table shows that the number of selected variables varies between 25 and 158 depending on the LASSO estimator used, with a deviance ratio in the range of 0.107 and 0.099. Our probit model includes 153 variables and produces for the testing sample a deviance ratio equal to 0.113. The fact that this value is close to the deviance ratio of the LASSO estimators shows that the predictive ability of our hand-curated set of variables is comparable to the ones chosen by the LASSO algorithm.[17]

While the previous test shows that we are using all the relevant information in the data set, a limitation is that we are conditioning on available data previous to the audit (that is if the audit is at time $t$, we condition on data up to time $t-1$). It may be that the audit at time $t$ relies on new soft information generated at time $t$ that is observed by the authorities but not by us. The problem with this is that we cannot condition on outcomes at time $t$ or after at the individual level, since these outcomes are potentially affected by the audit itself. This motivates the following balance test. We identify a set of variables at the province and sector level that may be correlated with the potential soft information available to the authorities at the time of the audit. This includes the average levels and growth rates of income from $t$ to $t+1$, VAT taxable turnover, operating costs, the ratio between operating costs and the net value of production, all at $t+1$. We also include the share of evasion at $t+1$, based on future audits in the same province and business sector at $t+1$. We use a fine partition of the sample (110 distinct geographic provinces and 21 business sectors), where each cell of this partition contains a median of about 400 tax returns.[18] We then test

[15] That is, we use post-estimation LASSO to make sure we do not form predictions using penalized estimates of coefficients. See for details Hastie et al. (2015).

[16] The deviance generalizes the residual sum of squares of the linear model and it is commonly used as a measure of fit of generalized linear models (Hastie et al., 2015). The deviance ratio is a generalization of the $R$-squared statistics, indicating the percentage of deviance explained by the model with respect to a model including only the intercept. Since the latter is always bounded between 0 and 1, it allows the comparison of alternative models.

[17] The randomness in the selection process explains a low deviance ratio. As we will mention later on, the IRA can review any of the tax filings over the past five years. Hence, each year, the population of tax filings at risk of being audited comprises all tax filings up to five years old that have not been audited in the previous years (about 72 million filings in our sample). The share of audits over the population at risk is 0.54% (388,513 audits divided by 72 million cases).

[18] The indicators listed above are informative of taxpayers' tax compliance behavior and ability to pay back the debt in a specific location and business sector. For example, a higher growth rate in income from a year to the next may signal lower propensity to evade taxes, a higher level or growth rate in revenues, VAT taxable turnover and operational costs may signal firms in expansion; a higher growth of the ratio

whether the expected values of these variables are systematically different in the population of audited and non audited taxpayers, conditioning (and unconditioning) on the observable variables. We should see that unconditioning, the means are naturally different in audited vs. non-audited taxpayers. Conditioning on the observables, however, we expect the conditional expectations to be independent of the audit for those generated by the IRA, and to be still dependent on the audit for the GdF — consistent with the fact that the IRA relies on the hard information in the Tax Registry, while the GdF also uses soft information).

Table 3 reports the OLS estimates of regression models where the levels at $t + 1$ of the measures of economic performance are regressed on a dummy variable taking value 1 if the taxpayer is audited at $t$ and zero otherwise.[19] Regressions in panels A and B are shown for all audits, whether initiated by the IRA or the GdF, first unconditionally (panel A) and then conditioning on the information available in the Tax Registry archives when audits are decided (panel B). With no controls, the conditional expectation of the measures of future performance at the province/sector level (panel A) depends on the audit. For example, the audit dummy at $t$ is significant for the level of income, total taxable revenues, VAT taxable turnover and operating costs (both in absolute and relative terms) at $t + 1$. The estimated coefficient of the audit variable loses statistical significance in most but not all the regressions when conditioning on data available to the tax authorities at the time of deciding audits (panel B). Decomposing the audits between those chosen by the IRA and by the GdF explains why the audit remains significant after conditioning on observables. When we exclusively focus on audits chosen by the GdF, we observe that the audit dummy is significant for future levels of taxable revenues, VAT taxable turnover and operating costs at the 5% level (panel C) even after controlling for hard information. This reflects the soft information gained on the ground by this police force.[20] When instead we consider only audits chosen by the IRA and we control for the hard data available at $t$, conditional expectations

between operational costs and the net value of production may signal distressed firms. A higher share of evasion may signal opportunities to hide income. There is ample casual evidence that the GdF uses "soft" information gained on the field to infer compliance habits of categories of taxpayers. These insights often translate in "campaigns" to audit specific professions (such as dentists and funeral homes) and in specific locations. For examples of campaigns on dentists and orthodontists, see Repubblica (1995) and Il Giornale Trentino (2014). For examples on funeral homes, see La Nuova Sardegna (2011) and Repubblica (2019). For examples of campaigns targeting specific regions, see Corriere della Sera (2012) reporting an auditing "blitz" in the popular ski resort of Cortina d'Ampezzo.

[19] The OLS estimates of regression models using growth rates rather than levels as dependent variables show a similar evidence. They are available upon requests.

[20] This evidence suggests that the soft information available to the GdF allows them to direct audits towards firms with more opportunities to hide larger sums of income (as signaled by correlation with evasion in own sector, and with the level of sales, VAT taxable turnover and operational costs) and that are relatively healthy (as signalled by lower ratios of operational costs over the value of production), and thus more likely to pay back any due taxes and fines.

of the economic variables at $t + 1$ are independent from the audit variable (panel D). The control function absorbs all the information in the audit decision of the IRA, apart from a residual random component. Given the institutional design of tax revenue and enforcement agencies discussed in Section 2.1, and the very low audit rate, this quasi-randomness of the audits decision by the IRA conditioning on the data in the Tax Registry should not be surprising. The rest of our study thus focuses on the 97% audits selected by the IRA and is based on regression models conditioning on the observed audit information.[21]

# 3 Tax advisors: who they are and what do they do

**Institutional setting.** Most TAs in Italy belong to professional orders, the largest of which is the order of the "Dottori Commercialisti."[22] These professional orders are regulated by law, similar to lawyers and civil engineers. Access to them is subject to stringent requirements: for the Commercialisti, it requires a university degree, a training period of 18 months working with a member of the order, and an admission exam. Besides regulating the behavior of the members, the law establishes *National Professional Boards* (NPB) comprised of elected members. The NPBs oversee the respective codes of conduct regulating how professionals interact among themselves, their clients and the tax authorities.

TAs have a uniquely important position for small businesses in Italy since they typically provide a wide variety of key services. The "Dottori Commercialisti", for example, besides filing tax returns keep the books of their clients (in 96% of the cases), provide management consulting (30%) and wealth management services (8.7%).[23] These activities give them a privileged position in terms of the knowledge they accumulate on the businesses of their clients. Because of this, they tend to establish a fiduciary relationship with their clients: Articles 200 and 351 of criminal procedure code and article 662 of criminal code explicitly regulates professional secrecy between the TA and client.

The use of a TA is not mandatory in Italy, but it is extremely common. In our sample

[21] Because the taxpayers audited by the GdF constitute 3% of the sample, the summary statistics reported in Table 1 remain roughly unchanged when excluding those audits from our sample. In Table S3 we report the statistics on the filings audited by the IRA and the GdF separately. The major difference is that audits by GdF target filings with average income 35% higher than the average income of filings audited by IRA. The share of filings with positive evasion and the average share of evasion detected is similar between audits by the two agencies, but the income evaded in case of positive evasion is much higher for audits by the GdF (2.6 times the one of audits by IRA). Finally, the duration of an audit is shorter for GdF than for IRA (75 days rather than 111).

[22] Depending on their professional training, TAs belong to three professional orders each one regulated by a specific law: the "Ordine dei Dottori Commercialisti e degli Esperti Contabili", the "Ordine dei Consulenti del Lavoro" and the "Registro dei Revisori Legali".

[23] See Di Nardo (2012)

of sole proprietorship, 97.5% of the taxpayers are served by a TA.

**Diffusion of TAs' services.** Table 4 shows summary statistics on the accountants in our data. Overall there are 107,069 TAs serving the 4.7 million taxpayers in our sample; a TA serves 31 sole proprietorship taxpayers on average, with large dispersion around the mean.[24] Accountants tend to serve taxpayers in their geographical proximity: in our sample an accountant has 62% of the customers in the same municipality and most (90%) in the same province. Over the sample period, we observe entries of new accountants and exits of existing ones. The average annual entry rate is 5.1% and the exit rate is 3.7%. Interestingly, while taxpayers tend to have long term relations with their TA, some do switch, sometimes as a consequence of closure of their accountant. Table 1, panel A shows that, on average, 7% of the taxpayers switch accountant over a year. Over all the sample period 18% of the taxpayers in the sample moved to a new TA, with one-third of such moves following the closure of the accountant. In Section 5.2 we rely on movers to test for sorting between taxpayers and accountants based on their propensity to evade taxes.

**Specialization and heterogeneity.** In Figures 1 and 2 we show the distribution of accountants by characteristics of their clients. In Figure 1, we study whether accountants specialize in clients of a particular sector. It shows the distribution of the number of sectors spanned by the TA clients, using a 2-digit (panel A) or a finer 5-digit sector code (ATECO classification, equivalent to the EU NACE; panel B). To avoid a mechanical tendency to show concentration, distributions are reported for TAs with at least 10 clients. Figure 1 reveals that when using the 2-digit classification, essentially no TA has all clients in one sector; almost all accountants (93%) serve clients in more than four 2-digit sectors, with a median of 8. The use of a more disaggregated sector definition (panel B) confirms a large dispersion in the clients sectoral composition: 96% of accountants have clients in at least ten different sectors. These figures are remarkable given that the average number of clients per TA is 31. As a further test of specialization we compare the empirical distribution in Figure 1 to a simulated distribution obtained by randomly assigning clients to TAs within the same province (shown by the dashed histograms in Figure 1). The empirical distribution almost fully overlaps the random assignment distribution: the Bhattacharyya coefficient of distribution overlap is 94% when considering the 2 digit sectors and 96% for the 5 digit sectors. All in all, since sole-proprietorship in our sample are small, unsophisticated business, unlikely larger sophisticated businesses they tend to demand relatively similar accounting

[24] The customer base of a TA is larger than this figure as they serve also incorporated firms as well as individual taxpayers.

services across sectors, which does not require TA sectoral specialization.[25] In Figure 2, we investigate other possible dimensions of specialization: the income of the clients, their age and that of the business. The figure shows clearly that accountants are highly heterogenous along all of these dimensions: we find no evidence that TAs specialize in serving taxpayers with low or high income, young or old entrepreneurs, or young or old firms.[26] Figure 3 provides descriptive information on heterogeneity across accountants along two additional dimensions: the share of customers that are audited (panel A) and the share of customers that are found to evade taxes, conditional on having at least two audited clients (panel B). Panel A reveals that around 40% of the accountants have no customer audited. The rest of the distribution shows marked heterogeneity across accountants in the fraction of their customers audited. The empirical distribution of the share of evaders among the audited customers of each accountant (panel B) also shows substantial heterogeneity and a long tail to the right: a few accountants have very large shares of evaders. The share of accountants with more than one-fourth of evaders among their audited clients is 89.9%, while few accountants with at least two audited clients have no evader among their audited clients.

**Size, turnout and profitability.** As reported in official statistics, most TAs operate as self-employed individuals (87% of the total in 2018). The remaining 13% are organized in associated offices of various types, where TAs share administrative costs and bill clients in a centralized way.[27] In associated offices, however, professionals are not necessarily employees: they maintain direct professional and legal responsibility for their clients as far as obedience to the code of conduct and the regulation is concerned. When organized in associate offices, TAs are rather small businesses: only 12.8% employ more than 10 people (including the TA, any other associate, trainees and employees of any type); 61.1% of the offices employ fewer than 5 people (Di Nardo, 2012).

When they operate as self-employed individuals, TAs file their own income as such. We are able to match self-employed TAs with their own tax filings, as well as the tax audits they have been subjected to. Accordingly, for matched TAs we can obtain measures of their turnout and profitability as well as of their personal filed income. Additionally, we observe whether and when they receive an audit and the (detected) TA personal evasion, if

[25] We will directly investigate the importance of TA sectoral specialization for our results in Section 4 (Table 7).

[26] In analogy with Figure 1, we report the counterfactual distribution of clients' characteristics under random TA assignment in the dashed histograms. The Bhattacharyya coefficients of overlap between the empirical and the random distribution are 93%, 92% and 93% for clients' income, age and experience, respectively.

[27] Official statistics on active firms by sector and legal form (ASIA Business Register of active enterprises, National Institute of Statistics data-warehouse, 2018).

any. We match more than 70% of the TAs in our sample with their tax filings. Matched TAs serve more than 60% of the taxpayers in our sample.[28] When compared, we do not find economically important differences between the characteristics of taxpayers served by matched and unmatched TAs (see Table S4 in the Online Appendix). The only notable difference is that unmatched TAs have more clients than matched ones (52 compared to 28), a reflection of the fact that, as mentioned, unmatched TAs work primarily in associated offices. This makes us confident that we can draw general conclusions when using data on tax filings of matched TAs, and their clients, to characterize TAs' heterogeneity.

Table 4, panel B provides summary statistics on the matched TAs' characteristics in our data. The majority of the accountants in our data are trained as "Dottori Commercialisti" (45%) or as "Ragionieri and Periti Commerciali" (31%). TAs are heterogeneous in terms of the size of their business as measured by filed income: on average they report about 42,500 euros per year, but the cross section ranges from 5,500 euros at the 10th percentile to 89,500 euros at the 90th percentile. The distribution is highly skewed to the right. There is a substantial heterogeneity in terms of profitability (measured by the ratio between the income of the TAs and the turnover of their business): the average is 0.33, smaller than the median (0.51) and the dispersion is large (the standard deviation is 0.46, the 90th-10th percentile difference is 0.67). TAs vary also substantially in terms of experience measured by the number of years in activity. Average experience is 18 years, with a standard deviation of 8.7 years.

**Tax evasion by TAs.** An important factor emerging from Table 4 is that TAs themselves are not completely averse to tax evasion, and show heterogeneity in rate of evasion. A distinctive feature of our data set is that we observe the outcome of audits on their tax filings, and thus whether they have been found evading**.** The median evasion after an audit on a TA is rather small, 1,996 euros; the mean evasion, however, is much larger: 18,760 euros. The share of evaders among the audited is 59.4%, smaller but in the same ballpark as the share of evaders among audited clients (66%, Table 1, panel B). The heterogeneity in the propensity to evade of the TAs is a key characteristics that we argue drives the market for intermediated tax evasion, and we use to test for clients sorting toward TAs with different tolerance for evasion.

[28]Unmatched TAs are TAs who do not file income as self-employed either because they work as employees in associated offices (80% of the unmatched) or because their main income source is as an employee and the TA activity is an aside (e.g. teachers with a secondary job as a TA). In the latter case TAs file as employees and report their income from the TA activities as one item in their tax filing (this group accounts for 20% of the unmatched). Matched and un-matched TAs can be identified from the type of tax identifier they report (i.e. fiscal code or VAT number).

**Quality of service.** Evasion by the TA and by his/her clients may reflect heterogenous *tolerance for evasion* for the TA and his/her clients, or may reflect heterogeneous TAs *quality*. Lower quality TAs may be more likely to make mistakes when filling in their clients' tax documents or even their own. Our data allow us to address this key issue, and show that clients' evasion cannot be simply a reflection of mistakes made by lower quality TAs. For each audited taxpayer we define a "misreport" as an instance in which there is a discrepancy between the reported income and the income assessed by the tax authority after an audit. Figure 4 shows the distribution of these misreports. While negative misreports (assessed income smaller than filed income) do occur, most misreports are positive, i.e. underreporting of income.

Table 5 studies whether the characteristics of TAs filing tax returns with positive or negative misreports are systematically different. We regress TA characteristics on dummies for the type of misreporting. The table shows that the experience, profitability and sectoral specialization (as measured by the number of 2-digit sectors spanned by the clients) of the TA have the same average value for negative and positive misreporting. Yet, TAs who evade more in their own practice are more likely to misreport in one specific direction: under-reporting their clients' income. This suggests that tax evasion is not the reflection of mistakes due to heterogeneity in TAs capabilities or sectoral specialization; it stems from heterogeneity in the evasion propensity of TAs and their clients. The remaining sections of the paper explore this issue in depth.

# 4 Exploratory evidence

We start by presenting a set of correlations that suggest that TAs play a role in tax compliance. The scope of this exercise is to set the stage for the next sections where we test two implications of the TAs as tax evasion facilitators model that can explain these correlations. To this goal, we first show the correlation between the taxpayer own tax evasion and the evasion *at* the TA, measured by the average tax evasion of the TA's other clients over the same time period. We then show the correlation between the tax evasion of a taxpayer and the evasion *of* his/her TA, measured directly exploiting the assessment of audits on TAs personal filings. We want to stress that here we make no claim of causality. What we want to show is that the correlations survive a set of controls that, independently of an active role of the TAs as tax evasion facilitators, may mechanically produce a correlation between the evasion of one taxpayer and that of the other clients of his TA; the most natural reason could be TAs specialize across industries and industries differ in tax evasion.

**TA's clients tax evasion.** We use several specifications of the linear regression model:

$$e_{ijk} = \alpha E_j + \beta controls_{ijk} + \varepsilon_{ijk} \tag{1}$$

where $i$ denotes a tax filing, $j$ denotes the TA, $k$ the fiscal year when income is produced and $\varepsilon_{ijk}$ are i.i.d., mean 0 innovations. The variable $e_{ijk}$ is the share of income evaded by taxpayers $i$ in fiscal year $k$, that is the difference between the income assessed during the audit and the declared income. $E_j$ is the average share of income evaded by the other customers of accountant $j$. The averages are computed over the entire period excluding the evasion of taxpayer $i$ for each year $k$. The first two columns of Table 6 show the OLS estimation results when using an increasing number of control variables (labeled as $controls_{ijk}$ in model 1). The sample includes all TAs with at least one other audited client. Standard errors are clustered at the accountant level in all regressions. In column 1 we include the variables that guide the selection of tax filings during the audit process as controls (see Section 2.3).[29] Results reveal that married and younger entrepreneurs, as well as owners of larger and older firms evade smaller shares of income. As in Kleven et al. (2011), the gender of the taxpayers has no predictive power once controls are included. As expected, being labelled as "congruent" and "coherent" by the tax authority correlates negatively and strongly with the share of evasion. The estimated value of $\alpha$, 0.118, is positive and highly statistically significant ($p$-value 0.000). It is also economically relevant: one standard deviation increase in the average share of evasion of the other clients of the accountant is associated with an higher share of own evasion of 2.3% (about 7% of the sample mean). Our baseline specification controls for location using a rich set of municipality dummies. Because Italy counts more than 8,000 municipalities, these very granular geographical controls ensure that the correlation between own evasion and the evasion of clients at the same accountant is not a reflection of omitted local factors. It is neither the reflection of unobserved geographical heterogeneity (e.g. in the propensity to pursue evasion) of the Italian tax authority as we include a full set of dummies for the IRA local offices. Finally, it cannot be the reflection of heterogeneity across economic sectors in (benefits or cost of) evasion coupled with TA sectoral specialization, because we control for business sector at a very fine level, and also find that TAs do not show sectoral specialization (Figure 1). In column 2 we add a rich set of TA characteristics including demographics (gender, age and marital status), professional training, experience, size, and profitability of their business as a measure of the quality of the services provided (see Section

[29] If there is only one other filing different from $i$ that is audited, the vector of controls contains the characteristics of that tax filing. If more than one tax filing is audited, we include the average of their characteristics.

2.3).[30] The estimated value of $\alpha$ is unaffected in size (0.116) and significance ($p$-value 0.000). In columns 3 and 4 we estimate the regression model 1 on a different sample and using a different estimation strategy. First, we exclude TAs with fewer than 50 customers, since average values may be misleading in very small groups. Second, we run our regression using a fractional probit specification (Papke and Wooldridge, 1996) since our dependent variable (the taxpayer's share of evasion) has a relatively large number of boundary values equal to 0 or 1. Results remain unchanged.[31] We read this correlation between a taxpayer's evasion and the average evasion of their TAs' other clients as suggestive of the TA's role in evasion.

**TA's clients tax evasion and their TA tax evasion.** To shed further light on this role we exploit the information resulting from audits on TAs. Specifically, we observe a TA's own evasion if they receive an audit. Accordingly, in column 5 we replace the TA's clients evasion rate with the TA's own evasion rate, holding all the other controls constant. Because we are relying on matched and audited TAs, this sample is much smaller. We find that the taxpayers evasion rate correlates positively and significantly with the evasion rate of their own TA. Furthermore, the relation between the taxpayer evasion and that of the TA is the same independently of the profitability or experience or sector specialization of the TA. To show this we divide the years of experience, profitability and specialization by sector of the TA into terciles, and interact the corresponding tercile dummies of each variable with the TA evasion. We then estimate our model 1 and compute an $F$ test for the null that correlation size between own and TA evasion is the same across terciles. Table 7 shows the results. In all cases we cannot reject the null hypothesis: the TA's own evasion predicts the evasion of their clients as strongly, independently of measured profitability, experience or sectoral specialization. In columns 4 and 5 of Table 7 we further show that the correlation does not vary with the number of clients served by the accountant and his/her professional training.

The evidence so far suggests that clients of the same TA tend to have similar evasion propensities, and that the TA's own evasion propensity, not other dimensions of heterogeneity

[30] Because these detailed characteristics are only observed in the sample of accountants matched with their own tax filing (the matched sample described in Section 3), they are multiplied by a dummy with value 1 for matched taxpayers and zero otherwise, with the dummy as an additional control. Results are robust when adding as additional characteristics the share of clients by age, sector, size, and the share of evaders on audited clients until the year $t$ in which filing $i$ is audited.

[31] To account for the fact that evasion is observed for the selected sample of audited taxpayers, we have also estimated the baseline regression using a two-stage Heckman regression model. Identification is achieved by relying on the non linearity of the probit model. The estimated $\alpha$ is 0.154, similar to that in the baseline regression and highly statistically significant. In addition, the evidence remains qualitatively unchanged if we use the entire sample and predict evasion using a machine learning algorithm to obtain a measure of expected evasion for each taxpayer. We extract a random 5% sample and run a LASSO model to select the Tax Registry variables that best predict evasion, where evasion is set to 0 for non-audited taxpayers. We use the selected variables and a linear specification to predict expected evasion for the remaining sample.

across TAs, contributes to this correlation. The latter is unlikely to be the reflection of omitted taxpayer, TAs or local factors since we account for these characteristics with an extremely rich control function. As an alternative test we follow Altonji et al. (2005) idea to assess whether the correlation may be driven by unobservables: if the later matter we should find that inserting finer and finer sector controls has relevant effects on the size of the correlation between own and peers evasion. We find this is not the case. For instance, controlling for 21 sector fixed effects instead of more than 1,000 five digits sector dummies leaves the coefficient of interest almost unaffected. The same is true if we jointly modify the geographical and sector controls: we find very little change in the slope of the relation between own and peer evasion when controls are made finer and finer. While not a proof, this suggests that that unobserved confounds are unlikely to be driving the correlation.

To further establish that the correlation documented in Table 6 arises because of a unique role played by the own TA, we estimate placebo regressions replacing $E_j$ in model 1 with the average share of evasion of the clients of a *different* but *similar* TA, one located near the taxpayer. We define similarity through a $k$-means clustering algorithm, assigning TAs in the same province to one of 20 groups based on observable characteristics of their clients. We run 1,000 regressions, each time randomly reassigning each taxpayer to a new accountant in the same cluster. Figure 5 shows the distribution of the estimated $\alpha$ for these placebos, as well as the distribution of the $t$-statistic of the null $\alpha = 0$. The spillover parameter is significantly different from 0 only in 4.7% of the cases. The actual estimate from Table 6 is instead sizable and more than ten times the maximum value of the estimates from placebo regressions. The conclusion is clear: the average share of evasion at accountants other than one's own bears no relation with own share of evasion except by chance. The correlation only arises when taxpayers share the same TA.[32]

# 5 The social role of tax advisors

TAs are not just middlemen between the tax authority and individual taxpayers, they instead play a role between the tax authority and the entire set of clients that they serve. From each interaction the TA gathers information on the tax authority's auditing strategy, learning

[32] We cluster accountants considering a large number of variables, including the number of clients, the number of different provinces of clients' residence, the fraction of women and married clients, the fraction of clients in each 2-digit sector, and averages of clients' age, firm size, years of activity, coherence, and congruence. Placebo tests exhibit absence of correlation also if we consider TA similarity along specific TA characteristics, such as sector and number of clients clients. Figure S1 in the Online Appendix reports the distributions of placebo coefficients obtained by reassigning taxpayers to other TAs in the: i) same province and with at least one client in the same 2-digit sector; or ii) same province and in the same decile of the size distribution.

more about what the tax authority audits and how. This is very valuable information, since the auditing strategy of a tax authority is one of the most guarded secrets. The information provides a reliable perspective on the cost-benefit trade off from evasion, perspective that the TAs can use to inform their remaining clients. For any other third party (e.g. a business association) it would be very difficult to elicit this information from the taxpayers: for privacy reasons and to avoid sharing with competitors valuable information, taxpayers would be reluctant to disclose details about their interactions with the tax authority. Taxpayers, moreover, gain from their TA's aggregate information, but have little incentives to contribute. TAs can naturally gather the information through their activities (e.g. book keeping and paycheck management), and then discretely repackage and incorporate it in their services.[33]

In the Appendix, we present a simple theoretical framework to model this activity and show that it can generate the sort of social spillovers described in the previous section. In addition to rationalizing the spillovers, the model also highlights additional predictions that we test empirically.

Two hypotheses emerge from the theoretical model. The first is the *information hub hypothesis* described above. We study this hypothesis in Section 5.1, where we show that an audit on a taxpayer generates important spillovers on their TA's other clients, and that the spillover passes through the TA. The panel structure and the size of our data are key to identify this effect, since we observe the response to an audit on the taxpayer's social network (in particular, the other clients) and the behavior in the social network *before and after* s/he joins a given TA.

The second is the *sorting for evasion services hypothesis* that can be seen as a corollary and a confirmation of the first. To the extent that TAs are providing services that facilitate tax avoidance and evasion, and to the extent that taxpayers and TAs are heterogeneous in evasion propensity, we should see sorting and a segmented market for intermediated evasion: taxpayers who are more willing or lenient to evade seek accountants that facilitate these activities; taxpayers who are not interested in tax evasion seek TAs with other "qualities". We turn to this prediction in Section 5.2, where we study whether taxpayers sort themselves on evasion propensity. Again, here the richness of our data is key, since we can directly observe a measure of the TAs' openness to evasion: their own evasion.

[33]The TA does not need to violate professional secrecy and/or disclose any sensitive information on individual taxpayers to other clients when sharing aggregate information. S/he instead only needs to formulate a professional recommendation on the risk of audit by taking into account the information on audits observed on other her/his clients.

## 5.1 Tax accountants as information hubs

### 5.1.1 Information spillovers

**Basic findings.** To study if TAs play any active role in diffusing auditing information among her/his clients, we study if an audit on a client at $t$ affects the reported income at $t+1$ of other clients. Tax audits are private, so we should expect no effect absent network effects.[34]

In Italy, income of year $t$ is reported from July 1st to September 30th of $t+1$. Accordingly, we refer to the income reported at time $t+1$ as the income reported between July 1st of $t+1$ and September 30th of $t+1$. We refer to the fiscal year $t$ as the period in between reporting time, so running from October 1st of $t$ to June 30th of $t+1$. In the analysis, we study the effect of an audit received during fiscal year $t-1$ on the income produced in fiscal year $t$ and filed at $t+1$. Audits concern income filed in at most the 5 previous years, i.e. years $t-1-j$ for $j = [1, ..., 5]$.

The first column of Table 8 shows the estimation of a simple regression of log taxable income filed at time $t+1$ where the only audit variable is an indicator equal to 1 if at least one of the clients of the same accountant was audited at $t-1$ (while excluding the taxpayer in question), labeled as *peer audit*.[35] In the second column, we include an indicator for whether the taxpayer was audited at $t-1$, labeled as *own audit*. All regression models include taxpayer fixed effects; effects on reported income at $t+1$ are relative to average taxpayer reporting. We are interested in the difference between a taxpayer's compliance behavior if the peers are treated with an audit, relative to no audit among peers. We also control for a set of time-varying taxpayer and accountant observables (marital status, age, size of the business, years of activity, accountant's clientele size and geographical coverage) at the time the audit is received, as well as time fixed effects. We include both the audit control variables related to a taxpayer's own tax filing being audited, and the average of these controls for the audited tax filings of peers.

The results in column 1 show that the effect of audits on peers is positive and highly statistically significant. When an indicator for whether a taxpayer was audited at $t$ is added in the regression (column 2), the peer audit effect is somewhat smaller in magnitude but remains large and precisely estimated. Having at least one other client of the same accountant audited triggers a 1.3% increase in filed income in the following year. If a taxpayer was

[34] We focus here on the effect of an audit on the other clients' reported income. In the Online Appendix Tables S1 and S2, we extend the analysis to show that an audit on a client influences the behavior of other clients in other ways as well.

[35] The results remain qualitatively unchanged if we use the precise number or share of clients audited at $t-1$, including the taxpayer in question, which is part of the set of signals observed by the TA.

audited at $t$, then in the subsequent year s/he reports a 7.5% higher income than average to the tax authority.[36]

**Robustness and placebo.** In column 3, we exploit the timing of the audits to construct a placebo test for the results presented above. Since audits in fiscal year $t-1$ do not belong to the information set of the taxpayer and of his/her TA when filing at $t-2$ , we replace the left hand side variable with the income reported at $t-2$. Consistent with the idea that peer audits, as well as own audits, induce taxpayers to change reporting behavior (because they revise their priors about the IRA policy), audits that are not yet in their information set have no effect on reported income: the peer and own audit coefficients are precisely estimated to be zero.

In the Online Appendix, we also show that the results in Table 8 are robust to how we measure the outcome variable and the nature of the audit. With respect to how we measure the outcome, past year peer audits and own audits both predict significantly current year tax compliance when measured by an indicator for reporting positive income, or reporting income above the 25th, the 50th or the 75th percentile (Table S5 in the Online Appendix). We note that two types of audits are possible: desk audits (67% of cases), or on site visits (33% of cases). Type of audit for a peer does not appear to matter for the spillover effect. As Table S6 in the Online Appendix shows, interacting the peer audit indicator with a bivariate variable equal to 1 if the audit is in person has no effect. This suggests that what matters is the presence of audits, and not their type. This may reflect the fact that the information on the type of audit is not considered relevant by the TA. On the other hand, the effect of own audits on compliance is significantly stronger — both economically and statistically — if the audit entails a visit. One reason may be that own in-person audits are remembered more vividly by the taxpayer, but this direct experience effect is not information that is passed by the TA to the peers.[37] Additional evidence that the TAs and their clients react to audits mainly because they learn about strategy of the tax authority is provided in the Online Appendix, Table S7. This table shows that an audit on a tax filing does not change the probability that other tax filings of clients of the same TA are audited in the following year, even for those in the same business sector or municipality of the audited filing. This suggests that previous year audits on filings of other clients do not mechanically affect the chance of being audited. Still, they provide useful information and affect future declared

[36]Incidentally, in unreported regressions we find that an audit to a TA customer also affects positively the reported income of the TA himself.

[37]The fact that in person visits have stronger effect than desk audits is in line with the evidence from the field experiment presented in Boning et al. (2020), in which US firms are treated by the IRS either with a letter or by a visit. Firms treated with a visit remit more taxes.

income because, in the presence of uncertainty on the audit probability, they are used by the TAs to update their beliefs on the audit probabilities of their respective clients.

### 5.1.2 Robustness to alternative social networks

An alternative story that may be consistent with the evidence discussed in the previous section is that the taxpayer learns about audits on their peers *directly* from the audited peers and not through the TA. This is possible if the taxpayers know each other, for example because they operate in the same area or business sector. We rule out this possibility by studying three direct implications of this hypothesis.

First, we focus on the clients of a TA. If peers interact directly among themselves, we should find stronger spillover effects between similar taxpayers, and weak or no effects between clients who operate in different cities and/or business sectors.

Second, we look at interactions between similar non-peers, i.e. taxpayers with similar characteristics who are served by different TAs. If informational spillovers are not channeled through the TA, we should find spillovers effects for audits on taxpayers that have similar characteristics even if they are *not* clients of the same accountant.

Third, we allow for other unobserved social networks among the clients of a TA (for example, the taxpayers are friends, former classmates, relatives). If there are such social links between the clients of the same TA, we should find the spillover to be present independently of whether taxpayers are served by the same TA. To show that this is not the case, we can exploit the size and longitudinal structure of our data: since we do observe behavior over time (before and after a taxpayers joins a given TA), we can measure if spillovers are present even after a taxpayer leaves the accountant, or right before a taxpayer becomes customer.

**Spillovers and taxpayers characteristics.** In testing whether there are direct information spillovers between similar taxpayers, we face the methodological challenge of identifying the appropriate cluster of similar taxpayers (both inside the TA, i.e. the similar client peers; and outside of the TA, i.e. similar non-client peers). To this goal, similarly to the procedure adopted to cluster the TAs in Figure 5, we use a standard machine learning technique, defining "similarity" non-parametrically using a $k$-means algorithm that assigns each filing within each province to one out of 100 clusters depending on a subset of observable characteristics (sector of activity, gender, civil status, years of activity, number of employees, coherence and congruence). We then test whether direct interactions among similar taxpayers, rather than TA intermediated interactions, drives the spillover from peer audits.

The results are contained in Table 9.[38] In the first column, we distinguish between audits

[38] The results are robust to changes in the number of clusters. See Table S8 in the Online Appendix.

on peers who belong to the same cluster as the taxpayer and on peers who belong to different clusters; that is we are holding constant the TA and varying the level of similarity of the audited clients. The peer audit effect is the same in both groups (0.013) and virtually equal to that in the baseline regression (Table 8). This is consistent with TAs sharing collected auditing information with their clients.[39] In the second column of Table 9 we add a variable capturing audits on clients of different accountants nearby but belonging to the same cluster as the taxpayer. If the spillover effect is driven by direct communication between taxpayers, it is natural to believe that we should find signs of spillovers between those who are more similar. Results show that this is not the case. The estimated effect on reported income of the audits on these similar taxpayers that are served by another TA is close to zero and not statistically significant. This suggests that it is the information disseminated by one's own TA that really matters.

**Unobserved social links between the customers of the same TA.** A further concern is that clients of the same TA may be socially connected through unobserved social links. Indeed, those social contacts may be the reason underlying the choice of a specific TA. To address this concern, we exploit the longitudinal dimension of our data set. We observe behavior and audits before and after a taxpayer selects a TA. If a taxpayer is connected to the clients of a TA (through links of friendships or family, for example), then s/he should also be socially connected to the clients before hiring or after leaving the TA. In column 3 of Table 9 we look at the effect of an audit at $t-1$ to a taxpayer who joins a TA at $t$ on the reported income of the customers of the new TA at $t$ (labelled as "peers in the future"). As it can be seen from the results, there is no information spillover among the clients of a TA when the audit is targeted on a taxpayer who is not yet a client but that will be a client the year after the audit. This evidence shows that the informational spillover does not flow through the social network that potentially leads the taxpayer to select the TA. Then, we look at the effect of an audit at $t-1$ to a taxpayer who joins a new TA at $t$ on the reported income of the customers of the old TA at $t$ (labelled as "peers in the past"). When a client is audited right after joining a new TA, s/he can still have personal relations with the old peers and inform them about the audit would affect their reported income. In column 4 of Table 9, we show that such spillover channel is absent.

[39]In Table S9 in the Online Appendix, we single out one of the most important drivers of diversity - business sector - and show that even in this case the estimated difference in the effect of peer audits in own sector and in different sectors is small and not statistically significant (column 1). The striking result (column 2) is that the effect of peer audit remains significant and of comparable magnitude even when the audited peers operate in different sectors and in different provinces. The evidence remains unchanged if we only focus the analysis on cases where there are no other audits of clients of different accountants in the same municipality (column 3), thus leaving the TA as the most probable driver of the spillover effect.

### 5.1.3 Dynamic effects

In Table 10 we investigate further the information mechanism by examining the persistence of the information effect. In the first column, we include the three lagged values of own audit at $t-1$, at $t-2$ and $t-3$ while controlling for other clients' audits at $t-1$. Interestingly, the effect of own audits is significant at all lags but the size decays over time, albeit slowly: the effect of a three-year old audit on current reported income is still 34% of the effect of a one-year old audit. The cumulative effect of an audit after three years is to increase reported income by 15.1% – almost twice as large as the one-year lagged effect. In this specification, the effect of an audit of other clients in the last year is significant and of the same size as in Table 8. In the second column, we also allow the audits of peers to affect reported income with lags of up to three years. The three lags are all positive and highly statistically significant. Importantly, once they enter together their size increases considerably. Perhaps most interestingly, the effect of the other audits observed by the TA on taxpayer reported income is larger for older audits. One potential explanation for this result is that information disseminates with lags. Another explanation is that details about the IRA policy are revealed as the audits unfold after they have already been notified. The variable for one-year lagged peer audits only captures the information about the IRA's notification of an audit to the taxpayers (and to the TA), while the two- and three-year old audits also reveal what the IRA investigates. This additional information allows the TAs to infer more about the IRA auditing policy. Both because estimated coefficients are larger and because several lags matter, the cumulative effect of the information spillover increases reported income by 8.6 percent — about 57% of the cumulative direct effect of an audit. When we include audit policy controls for both own and peer audits and for all the different lags, the magnitude of the estimated coefficients not surprisingly decreases (third column). The effects, however, remain statistically significant and follow the same patterns over time. The indirect cumulative effect is about 16% of the direct effect. This is a non-negligible effect, since the indirect effect is at work for the entire population.

In Section 6 we estimate the deterrence effect generated by the information spillover in terms of additional taxable income reported, showing that it should be accounted as an important factor when evaluating the efficacy of the auditing policy. Appendix B provides evidence that TAs do not share information on audits to their own clients with other TAs and discuss reasons for why they do not.

## 5.2 The market for tax-evasion facilitators

### 5.2.1 Sorting over services for tax evasion

When TAs are heterogeneous in their willingness to act as information hubs and the taxpayers have heterogeneous preferences for their services, we should expect taxpayers to sort themselves out on the basis to their willingness to evade taxation, or more broadly to engage in aggressive income reporting strategies, which may or may not be ruled as illegal evasion if an audit actually occurs, in light of ambiguous tax rules. Only for brevity we label this type of sorting as sorting on tax evasion services and the corresponding market as market for intermediated tax evasion services.[40]

To study whether there is sorting in the market for tax evasion facilitators, we exploit the longitudinal structure of our data set, studying the choices of taxpayers who switch TAs.[41]

First, we look at the correlation between the average tax evasion of the clients of a TA before the move of one of her clients, and that of the clients of the TA reached after the move. Sorting implies that, upon moving, a client should match with a new TA whose clients have a similar average tax-compliance as the clients of the preceding TA. Figure 6, shows a bin-scatter of the share of evasion at the old and new accountants, after we partial out year, sector and municipality fixed effects as well as a set of movers' characteristics (gender and age of taxpayer, size and duration of the business, year of move). These controls are primarily meant to mitigate the possibility that evasion rates have a local and/or sectoral component. The figure shows the (non-parametric) relationship for the whole sample of movers (672,084 taxpayers for which we can compute the average evasion of the clients of the old and new accountant) and is unambiguously strongly positive.[42]

The size of our sample allows us to refine this evidence further by running our regressions at the individual level and focusing on taxpayers that switch, and were audited at least once *before* switching accountants. We can then measure the correlation between the tax evasion of the mover when s/he was served by the old accountant and the average tax evasion of the clients of the new TA before the move. The results of this regression analysis

[40] Hemel et al. (2021) show the existence of many gray areas where the law is ambiguous making the determination of the "true tax" debatable.

[41] See Proposition 2 in Section A in the Appendix. One may ask how evasion prone taxpayer obtain information on TAs willing to offer tax evasion services. One natural way this information is collected is among friends/relatives and acquaintances: homophily implies that tax evaders will have as friends other evaders that may be informed about the market for advisors. A recent survey shows that indeed search for TAs occurs mostly through word of mouth among friends and social relations (CENSIS-ENPACL, 2017).

[42] The evidence is confirmed in a formal regression analysis of the average evasion of the clients of the old accountant on the average evasion of the clients of the new accountant, controlling also for year of move fixed effects and the entire list of audit policy controls at the accountant level for both the old and the new accountants.

are reported in Table 11. To make sure that moves between accountants are not renames of the old accountant, we exclude switches to TAs that appear for the first time (14% of the closure events). Because we focus on audited switchers, the sample size shrinks to around 30 thousands taxpayers but remains large enough for reliable inference. In the first column, we measure the evasion of the receiving TA using the average evasion of other clients before the switch occurs. We add an extensive set of controls, including the fixed effects for the year of the move, the controls for the selection of the audited filings of the mover and of other taxpayers, and the characteristics of the old TA. The results show that the correlation between own evasion at the old accountant and average evasion of the clients at the new accountant *before* the switch is positive, highly statistically significant and very similar in size to the slope values underlying the relation in Figure 6.[43] In the second column, we add an interaction term between the evasion of clients of the new accountant and a binary variable with value 1 in case the old accountant is no longer active, to isolate cases of involuntary moves. Results show that the effect remains positive and significant. It is smaller in magnitude for the cases of involuntary moves. In the third column we look closely to whether the evidence on sorting holds irrespective of the motivation triggering the move. One may argue that the evidence of sorting may be due to a mechanical effect of mergers between TAs. We identify these cases by adopting the definition used in the matched employer-employee literature for firm acquisitions. Following Fink et al. (2010), we define as take-overs all the cases in which we observe the exit of the old TA and a large group of clients jointly moving from this TA to another TA. Namely, the large group is defined as at least 50 clients or at least 50% of the clients of a closing TA with at least 10 clients. This group includes about 4,500 cases (24% of the closure events). In addition, we identify closures caused by the retirement or the death of the TA — events that unavoidably force the taxpayer to re-match with a new, different TA. In column 3, we use the whole sample and interact the average evasion of other clients with the different (mutually exclusive) types of closures. We cannot reject the null that the degree of sorting is the same independently of the reason that triggered the move ($F$ test for the interaction terms: 0.35, $p$-value: 0.702). These various exercises suggest that there is a genuine evidence of assortative mating of taxpayers and TAs by propensity to evade.

**Robustness and placebos.** The results are robust to sample selection and estimation method, as shown by the estimates in column 4 where we only include large TAs (those with more than 50 clients), and in column 5 where we use a fractional probit (as we do for Table

[43]The evidence remains qualitatively unchanged if we run the analysis on the entire sample of taxpayers and use as a measure of evasion the predicted evasion described in footnote 31.

6). Finally, we run a placebo test to corroborate the evidence presented thus far. Using a similar procedure of Figure 5, we match a switcher with a new, randomly selected TA in the same cluster of the true new TA, then estimate the regression in the first column of Table 11 and record the coefficient on the average share of evasion at the previous accountant and its significance. Figure 7 shows the distribution of the estimated slope parameter and of the corresponding $t$-statistic. The figure shows that the estimates are small and centered around zero. The coefficients are statistically different from zero in less than 1% of the cases only. All placebo estimates are much smaller than the actual estimate in Table 11.

### 5.2.2 Robustness to other potential drivers of sorting

A potential threat to our interpretation of the evidence of sorting so far is that taxpayers sort themselves on the basis of other criteria that happen to be correlated with tax evasion. This concern is mitigated by the evidence presented in Section 3: TAs are not specialized in serving taxpayer in a specific business sector or other characteristics (such as by income level, new or established firms, young or old professionals). In this section, we further address this concern in two ways. First, we show that the results presented above are robust to controlling for measures of TA's sectoral specialization. Second, we show that the own evasion of the TA is a key driver of the observed sorting.

**Does specialization play a role in sorting?** To further investigate whether our results are driven by TA's specialization, we add measures of TA's specialization as additional controls in our baseline regression. In Table 12 we estimate the sorting regression controlling for whether the new TA has most of the clients in the same sector as the mover (column 1), or whether the modal sector of the clients of the new TA and old TA is the same (column 2). If sectoral specialization was correlated with evasion, and would thus be driving the result, we should observe a significant drop in the relation between the evasion at the new TA and that of the mover before the move. As the table shows, the coefficient (0.042) is the same as that estimated in the first column of Table 11, and the measures of TA sectoral specialization have no explanatory power. We find a similar result if we control for the new and old TAs having the same type of training background (column 3), as another potential dimension of specialization. In Table S10 in the Online Appendix we add measures of TA specialization along the dimensions considered in Figure 2. The results show no decrease in the estimated correlation, suggesting that TA specialization does not mediate the evidence of sorting.

**Sorting and the TAs' ethical standards.** To further strengthen the evidence that TA tolerance for evasion drives sorting, we present estimates in Table 13 of the same specifications as in Table 11, but on the sample of TAs that were audited, using the TA own evasion detected in an audit instead of the average tax evasion of their clients. The sample is much smaller (1,634 observations) but the estimates speak clearly: when a taxpayer leaves a TA for a new one, s/he looks for a new TA that has the same propensity for evasion as the old TA. As the other columns show, the positive correlation is not driven by the TA sectoral specialization or their training. Interestingly, the slope coefficient is twice as large when evasion propensity of the new accountant is measured with his/her personal evasion instead of clients evasion, arguably because his/her personal evasion is a better measure of the TA type in terms of evasion propensity.

## 5.3 A broader view of the market for intermediated tax evasion

The evasion facilitation role of TAs goes beyond sharing information of frequency of audits. TAs gather information not only on audit probability but also on the detection probability and amount found in an IRA audit, revealing information on the IRA's ability to spot hidden income. Additionally, TAs can learn about the ability of the IRA to collect the amount of detected evasion if the taxpayer appeals. Because TAs intermediate between the taxpayers and the agency, as part of this process they entertain direct and frequent relations with the IRA local branches, which can prove valuable in assisting their clients if, following an audit, a tax controversy arises.

We now provide some evidence of this broader role of TAs. In Figure 8 we show that taxpayers sort on TAs not only on the basis of evasion propensity, but also on the measure of tax avoidance discussed in Section 4 and on the fraction of appeals set up by the TA. The latter measure captures the idea that the decision to appeal to a request of the IRA to pay for detected evasion after an audit reveals the chances of winning the appeal available to the TA, which is clearly valuable for a tax evasion inclined taxpayer. Looking at taxpayers moving to a new accountant, we find a strong positive correlation between clients of the old and new accountant in terms of both the average tax avoidance (panel A) and the average appeal rate after an audit (panel B).[44] This evidence suggests that in the market for intermediated tax evasion, taxpayers value TA advice on dimensions of the IRA policy that are relevant for their reporting decision and that go beyond the probability of an audit. We can shed some light on the nature of this advice and on how TAs can tailor the tax returns of their clients to help them evade. As mentioned in Section 2, a small portion of taxpayers in Italy does not

[44]Formal regressions confirm this evidence. As reported in Table S11 in the Online Appendix, the tax avoidance filed by a taxpayer changing accountant before the switch is positively related with the average tax avoidance of the clients of the new accountant before the switch. Similarly, the probability than a mover who has been audited appeals is positively correlated with the appeal rate of the clients of the receiving accountant.

hire a TA (around 3% of the universe of sole proprietor taxpayers). In Table S12 we study whether taxpayers with and without the advice of a TA differ systematically by comparing key financial variables, keeping constant income, sector of activity, municipality and size of the firm. We consider the amount of deductions (as measured by our tax avoidance index), expenses incurred from the normal day-to-day of running a business, the accumulated costs of fixed assets, and the amount of activities generating VAT. Results show that taxpayers with TAs on average claim higher deductions, higher costs, and a lower level of VAT, even controlling for income, sector, municipality and size of the firm. This evidence suggests that TAs help their clients restructure their financial accounts in a way such that they decrease their taxable income.

# 6 Policy implications

The results presented so far have important policy implications concerning the optimal design of the audit policy. Moreover, they are relevant for various current debates on tax evasion and enforcement, concerning the evaluation of the performance of tax agencies and the regulation of tax professionals.

With respect to the design of audit policies, two lessons emerge clearly. First, our analysis suggests that the IRA should keep track of the outcome of the audits on TAs or on their clients, targeting further audits on the clients of TAs who are found to evade on their own taxes and/or have clients evading taxes. This is a direct implication of the sorting documented in Section 5.2, implying that the evasion of the TA or of his/her audited clients predicts that of the non-audited clients. Second, our findings suggest that, after a successful audit of a client of a TA, audits should preferably target income preceding or contemporaneous to the audit, not for following years. The effects documented in Section 5.1 implies that taxpayers will respond to an audit on themselves or a peer with a temporary increase in reported income, thus making them less likely to be found evading in the future.

Table 14 shows that neither of these two policy lessons have been incorporated in Italy by the current IRA policy. To verify this, we estimated probit models of the probability that a filing receives an audit depending on previous audits on the peers and accountant. The first column shows that an audit on peers in the year before the filing does not decrease the probability of receiving an audit, not even when the audit detects positive evasion, therefore the indirect enforcement effect is not taken into account. In columns 2 and 3, we also find that the audit risk does not change after the assessment of positive evasion in the accountant's filings. Thus, our analysis shows concrete directions for cost-effective interventions to improve the audit selection design.

Of course, there is no guarantee that if the IRA changes policy and incorporates the lesson highlighted above, taxpayers will not respond by changing their behavior and make the policy changes less useful. Our observations on how to improve on existing policy given current behavior, however, is relevant for three reasons. First, the audit policy (and any change of it) is unobserved by the taxpayers, so even assuming full rationality and forward looking expectations, it would take a significant amount of time for taxpayers (and their TAs) to learn about the policy. Second, assuming full rationality and forward looking behavior seems implausible in this context, since some of the reactions shown by the empirical analysis probably have a behavioral component that would persist over time. Finally, and perhaps most importantly, our analysis does not imply that the IRA should update its policies using existing data only once: the policy should constantly monitor behavior and adjust over time in a dynamic an interactive fashion. The results presented above should be interpreted as an example of the existence of margins for improvement.

Our analysis has also implications for how audit policies are evaluated. The results on information spillovers of Section 5.1 imply that an audit does not only discourage future evasion of the audited taxpayer but, through the TAs, it also produces spillover effects on the reporting behavior of the other clients, and these effects persist over time. To get a better sense of the quantitative importance of the information channel on reported income, consider increasing the number of audits by one unit for each TA. Our estimates imply that the total cumulative direct effect, over three years, on the reported income of these taxpayers amounts to EUR 1,351 millions, and the information spillover effect amounts to EUR 711 millions — approximately 53% of the direct effect. The role of TAs as information hubs works as a multiplier for the deterrence effect, and it should be exploited by the IRA.[45]

This deterrence effect should be considered in the evaluation of the audit policy and the allocation of funds to the internal revenue agencies. Governments allocate money to audit activities based on how effective they are in contrasting evasion and maximizing revenues. But at least in Italy (see for example, Corte dei Conti, 2019), the spillover effects of the audit policy are not accounted for in the cost-benefit analysis, implying that audit activities are under-funded.

Finally, our paper contributes to a recent hot debate on whether TAs should be considered legally liable for their clients evasion, and more generally about the degree to which

[45] See Appendix C for details of these calculations. Alternatively, we can focus on the effect of audits on tax revenues accounting from TA spillovers. Audits affect tax revenues in two ways: a) mechanically when an audit detects evasion and the later is recovered; b) through the discussed deterrence effect that induced higher future reported income. A back of the envelope calculation for year 2012 results in an estimate of additional tax revenues from the audits of 656 million, of which 66% from the detected (and recovered) evasion and the rest from the deterrence effect, equally split between the direct and spillover effect through the TA. See Appendix C for details.

professionals should be responsible for the actions of their clients.[46] We provide evidence that TAs play an important role in facilitating tax evasions and document the existence of a market for intermediated tax evasion, where evasion-prone taxpayers match with evasion-tolerant facilitators. This evidence supports the idea that making tax professionals liable for their clients' evasion may discourage the market for intermediated tax evasion. In the US, thanks to a broad interpretation of the Section 7201 of the Internal Revenue Code, a tax practitioner can be found guilty to the same extent as the taxpayer who actually owes the taxes if s/he helps evading. A recent European Directive (DAC6, 2018/822) is also moving steps in this direction. It requires TAs to report to the police any information they may have about taxpayer's fraudulent behaviors/requests related to cross-border operations.[47] Our evidence suggests that the involvement of some TAs is systematic, thus supports the design of legal norms that acknowledge this role and find ways to make TAs more legally accountable.

# 7 Conclusions

Tax codes in advanced countries have become increasingly complex, creating scope for experts' advice. We argue that, depending on the role played, tax intermediaries can have profound effects on the nature of the relationship between tax authorities and taxpayers. TAs may help taxpayers take advantage of the complexity of tax rules and game tax authorities by offering taxpayer-specific counseling on how to minimize income reporting within or outside the boundaries of the tax code. This implies a market emerges where (some) accountants specialize in offering evasion advice to evasion-prone taxpayers. We find strong evidence that evasion-prone taxpayers match with evasion prone-TAs, implying that some accountants specialize as tax-evasion facilitators. A smart tax authority should then invest resources to learn the accountants' types, diverting attention from the taxpayers to their intermediaries, and audit with higher probability clients of more evasion-prone accountants. This breaks the direct link between the tax authority and the taxpayers assumed in the traditional literature on tax evasion and compliance (e.g. Allingham and Sandmo, 1972; Graetz et al., 1986). In these models, absent tax intermediaries, taxpayers comply only because they can be audited with some probability and punished if found non-compliant. With tax

[46]While the Italian law punishes tax preparers if they are found to contribute to the production of false invoices, in recent years the Italian Supreme Court of Cassation was repeatedly called to express its view on the punishability of the tax preparer when they act as "partner in crime" in the evasion of the client, and it has increasingly broadened the set of circumstances where the TA is legally liable for tax fraud.

[47]That is, to act as whistleblowers. The directive has been recently incorporated in the Italian code by law D. lgs. 100/2020, effective since August 2020.

intermediaries, taxpayers can also be disciplined by the audits of other clients of their own accountant. Accountants may act as information hubs: taxpayers can learn about the tax authority's policy because accountants can pool the audit experiences of many customers over many years and share this information with their clients. From the point of view of the taxpayer, this speeds up learning about the tax authority policy function, providing an additional incentive to rely on TAs. From the point of view of the tax authority, auditing one taxpayer can, through the information disseminated by the accountant, affect the compliance of the other clients. We find evidence that this is indeed the case. Reported income not only responds positively to a directly experienced audit but also to the audits of the other customers of one's own TA. The size and pattern of responses to the two types of audits is telling: taxpayers' response to own audits is strong on impact, but its effect is short-lived and decreases rapidly with time. The response to other clients' audits is milder on impact but persists unchanged over time. One interpretation is that own audits have much greater salience than others' audits, but salience vanishes as distance from the audit increases. On the other hand, at each point in time, accountants are much more likely than single taxpayers to observe an audit. Passing on this information to their clients increases audit salience. Accountants have the ability to keep track of all previous audits of their clients: information accumulates and becomes more precise as time lapses. Understanding the dynamic response to direct and indirect exposure to audits is both intriguing and of practical relevance to evaluate the effects of audit policies and improve their design. Our analysis moves a first step in this direction.

# References


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

Alm, J., K. M. Bloomquist, and M. McKee (2017). When You Know Your Neighbour Pays Taxes: Information, Peer Effects and Tax Compliance. *Fiscal Studies 38*(4), 587–613.

Alm, J., B. R. Jackson, and M. McKee (2009). Getting the Word Out: Enforcement Information Dissemination and Compliance Behavior. *Journal of Public Economics 93*(3-4), 392–402.

Altonji, J. G., T. E. Elder, and C. R. Taber (2005). Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools. *Journal of Political Economy 113*(1), 151–84.

Andreoni, J., B. Erard, and J. Feinstein (1998). Tax Compliance. *Journal of Economic Literature 36*(2), 818–860.

Beck, P. J. and W.-O. Jung (1989). The Role of Tax Preparers in Tax Compliance. *Policy Sciences 22*(2), 167–194.

Boning, W. C., J. Guyton, R. H. Hodge, and J. Slemrod (2020). Heard it Through the Grapevine: The Direct and Network Effects of a Tax Enforcement Field Experiment on Firms. *Journal of Public Economics 190.*

Braithwaite, J. (2005). Globalisation, Redistribution and Tax Avoidance. *Public Policy Research 12*(2), 85–92.

CENSIS-ENPACL (2017). Crescita e consolidamento nel futuro dei consulenti del lavoro.

Chetty, R., J. N. Friedman, and E. Saez (2013). Using Differences in Knowledge across Neighborhoods to Uncover the Impacts of the EITC on Earnings. *American Economic Review 103*(7), 2683–2721.

Chetty, R. and E. Saez (2013). Teaching the Tax Code: Earnings Responses to an Experiment with EITC Recipients. *American Economic Journal: Applied Economics 5*(1), 1–31.

Corriere della Sera (2012). Blitz di fine anno, l'ira di Cortina. Controlli fiscali con 80 agenti.

Corte dei Conti (2019). Relazione sul Rendiconto Generale dello Stato 2019.

D'Agosto, E., M. Manzo, S. Pisani, and F. M. D'Arcangelo (2017). The Effect of Audit Activity on Tax Declaration: Evidence on Small Businesses in Italy. *Public Finance Review 46*(1), 29–57.

DeGroot, M. H. (1970). *Optimal Statistical Decisions*. John Wiley and Sons.

Del Carpio, L. (2014). Are the Neighbors Cheating? Evidence from a Social Norm Experiment on Property Taxes in Peru. *mimeo, Princeton University*.

Di Nardo, T. (2012). *L'Indagine statistica nazionale 2012*. Fondazione Istituto di Ricerca dei Dottori Commercialisti e degli Esperti Contabili.

Erard, B. (1993). Taxation with Representation: an Analysis of the Role of Tax Practitioners in Tax Compliance. *Journal of Public Economics 52*(2), 163–197.

Fink, M., E. Kalkbrenner, A. Weber, and C. Zulehner (2010). Extracting Firm Information from Administrative Records: The ASSD Firm Panel. *Working Paper* (1004).

Fortin, B., G. Lacroix, and M.-C. Villeval (2007). Tax Evasion and Social Interactions. *Journal of Public Economics 91*, 2089–2112.

Galbiati, R. and G. Zanella (2012). The Tax Evasion Social Multiplier: Evidence from Italy. *Journal of Public Economics 96*(5), 485 – 494.

Garin, A., E. Jackson, and D. Koustas (2021). New Work or Changes in Reporting? A Framework for Measuring Self-Employment Trends in Administrative Data. *Working paper, London School of Economics*.

Graetz, M. J., J. F. Reinganum, and L. Wilde (1986). The Tax Compliance Game: toward an Interactive Theory of Law Enforcement. *Journal of Law Economics and Organization 2*(1), 1–32.

Harrington, B. (2016). *Capital without Borders: Wealth Managers and the One Percent*. Cambridge, MA: Harvard University Press.

Hastie, T., R. J. Tibshirani, and M. Wainwright (2015). *Statistical Learning with Sparsity: The Lasso and Generalizations*. Chapman and Hall/CRC.

Hemel, D. J., J. Holtzblatt, and S. Rosenthal (2021). The Tax Gap's Many Shades of Gray. *University of Chicago Coase-Sandor Institute for Law & Economics Research Paper 938*.

Il Giornale Trentino (2014, October, 28). I dentisti nel mirino della Guardia di Finanza.

Klepper, S., M. Mazur, and D. Nagin (1991). Expert Intermediaries and Legal Compliance: the Case of Tax Preparers. *Journal of Law and Economics 34*(1), 205–29.

Kleven, H. J., M. B. Knudsen, C. T. Kreiner, S. Pedersen, and E. Saez (2011). Unwilling or Unable to Cheat? Evidence from a Tax Audit Experiment in Denmark. *Econometrica 79*(3), 651–92.

La Nuova Sardegna (2011, October, 22). Maxi evasione fiscale, nei guai impresa di pompe funebri.

OECD (2016). Italy's Tax Administration: A Review of Institutional and Governance Aspects.

Papke, L. E. and J. M. Wooldridge (1996). Econometric Methods for Fractional Response Variables with an Application to 401(k) Plan Participation Rates. *Journal of Applied Econometrics 11*, 619–632.

Perez-Truglia, R. and U. Troiano (2018). Shaming Tax Delinquents. *Journal of Public Economics 167*, 120–137.

Phillips, M. D. (2014). Deterrence vs. Gamesmanship: Taxpayer Response to Targeted Audits and Endogenous Penalties. *Journal of Economic Behavior & Organization 100*, 81–98.

Pomeranz, D. (2015). No Taxation without Information: Deterrence and Self-Enforcement in the Value Added Tax. *American Economic Review 105*(8), 2539–69.

Raskolnikov, A. (2007). The Cost of Norms: Tax Effects of Tacit Understandings. *University of Chicago Law Review 74*(2).

Reinganum, J. F. and L. L. Wilde (1991). Equilibrium Enforcement and Compliance in the Presence of Tax Practitioners. *Journal of Law, Economics and Organization 7*(1), 163–181.

Repubblica (1995, January, 31). Battaglia tra Fisco e dentisti.

Repubblica (2019, August, 22). Fisco, tasse evase per 6 milioni: indagine su ditte pompe funebri a Napoli.

Rincke, J. and C. Traxler (2011). Enforcement Spillovers. *The Review of Economics and Statistics 93*(4), 1224–1234.

Saez, E. (2010). Do Taxpayers Bunch at Kink Points? *American Economic Journal: Economic Policy 2*(3), 180–212.

Scotchmer, S. (1989). Who Profits from Taxpayer Confusion? *Economics Letters 29*(1), 49–55.

Slemrod, J. (1989). The Return To Tax Simplification: an Econometric Analysis. *Public Finance Quarterly 17*(1), 3–27.

Smith, K. W. and K. A. Kinsey (1987). Understanding Taxpaying Behavior: a Conceptual Framework with Implications for Research. *Law and Society Review 21*(4), 639–663.

Yitzhaki, S. (1974). Income Tax Evasion: A Theoretical Analysis. *Journal of Public Economics 3*(2), 201–202.

Zwick, E. (2021). The Costs of Corporate Tax Complexity. *American Economic Journal: Economic Policy 13*(2), 467–500.

# APPENDIX

## A A Model of Tax Professionals and Their Clients

We present here a simple theoretical framework to study the relationship between the tax authority, the tax accountants (TAs) and their clients. The standard game theoretic model of tax evasion in which taxpayers act independently of each other is augmented to allow for the presence of TAs who collect information on the tax authority's strategy from its activity and coordinates the evasion rates of his/her customers.[48] The goal is to rationalize the social spillovers among the clients of the same TA and thus to formally derive the testable hypotheses that are studied in the paper. In the model, accountants optimally choose the evasion of the taxpayers conditional on the information that, thanks to their role, they can aggregate by observing several realizations of tax audits; and they use it to anticipate the IRA auditing probabilities. In turn, the tax authority chooses these probabilities optimally to maximize tax revenues net of auditing cost. The model predicts a positive relationship between individual tax evasion and both the average tax evasion of the other clients of the TA, and the average observed tax evasion (i.e. the average evasion observed by the IRA with the audits). We highlight two mechanisms generating this correlation: self-selection of taxpayers into accountants with heterogeneous attitudes about tax evasion, and informational externalities generated in the TA's activities.

**Setup.** Assume there is a continuum of taxpayers with mass one. Each taxpayer $i$ is associated to an income $Y_i$ unobserved by the tax authority, and can evade a share of income

[48]For traditional models of tax evasion in which taxpayers act interdependently see Allingham and Sandmo (1972) and Yitzhaki (1974). More recently, Phillips (2014) has presented a model of tax evasion in which the probability of an audit may depend on the share of a taxpayer's income that is matched by a third party and on the share of unmatched income that is self-reported by the taxpayer. Phillips argues that the (positive) elasticity of the probability of an audit on self reported income may explain high rates of compliance in the face of low expected penalties for evasion..

$e_i \in [0,1]$.[49] Taxpayer $i$'s utility of choosing a level of tax evasion $e_i$ with TA $j$ is assumed to be:

$$u(e_i, m_i) = [(1-p_i)(e_i Y_i) - p_i (T e_i Y_i)] - F_j (e_i Y_i) - C(e_i, Y_i, m_i). \tag{2}$$

where $p_i$ is the probability of being audited and discovered and $Te_iY_i$ is the cost of being audited with evasion $e_iY_i$. The term in brackets is the net expected benefit of the tax evasion; the second term, $F_j(e_iY_i)$, is the fee to be paid to the TA who prepares the tax returns, that may be a function of the level of evasion (in the remainder we assume all taxpayers need a TA).[50] The last term, $C(e_i, Y_i, m_i)$, is the "ethical" cost of violating the law, where $m_i$ is the taxpayer's type. We assume $C(e_i, Y_i, m_i)$ is increasing and convex in $e_i$ and that types with a lower $m_i$ are more "ethical": $C'_{m_i}(e_i, Y_i, m_i) < 0$ and $C''_{e_i m_i}(e_i, Y_i, m_i) < 0$, so that the lower is $m_i$, the more costly it is to violate the law and the higher is the marginal cost. We also assume that the elasticity of evasion with respect to income is lower than 1, i.e. $\eta < 1$.[51] This assumption is not necessary, but it seems the most plausible since, as we will see, it implies that the rate of evasion is not decreasing in the income of the taxpayer. In the following we use the functional form $C(e_i, Y_i, m_i) = \frac{(e_i)^2}{2m_i}(Y_i)^\eta$ for $\eta \leq 1$, which satisfies all these conditions.[52] The ethical standard $m_i$ is observed only by the taxpayer and has uniform distribution with support $[\underline{m}, \overline{m}]$. Taxpayers are however assumed to belong to one of $K$ observable categories: these may correspond to the general business activity and/or demographics, or both.[53] The income of a taxpayer $i$ in category $k$ is a random variable $Y_i$ with Pareto distribution with minimum support $\underline{Y}^k$ and shape $\rho_k$. so

$$f_k(Y_i) = \rho_k \cdot \frac{\left(\underline{Y}^k\right)^{\rho_k}}{(Y_i)^{\rho_k+1}} = P(Y_i; \underline{Y}^k, \rho_k)$$

for $Y \geq \underline{Y}^k$. The coefficients $Y_k$ and $\rho_k$ are assumed increasing in $k$, so categories with larger index have a larger expected income. The complete type of a taxpayer is $\theta_i = (m_i, Y_i, k)$.

We assume that there is a finite number $J$ of TAs with heterogeneous dispositions to allow their customers to violate the law. Specifically, the utility of an accountant $j$ who

[49] If a share $\tau$ of income is owed to the government and the taxpayer evades a share $\tilde{e}$ of the amount due, then $e = \tilde{e}\tau$.

[50] In our data set, only 2.5% of taxpayers choose to file without a TA.

[51] As we will see in greater detail below, the optimal evasion for the taxpayer can be written as $Am_iY_i^\eta$ where $A$ is a constant: so the elasticity with respect to income is $\epsilon_{e_iY_i} = \eta$, which is lower than 1 for $\eta < 1$.

[52] When $\eta = 1$ the marginal cost of evading is proportional to $Y_i$,which implies that the trade off for a taxpayer doe snot change with income, and the optimal rate of evasion is constant; when $\eta$<1 instead as income increases, the taxpayer likes to increase the rate of evasion. The case with $\eta > 1$is less interesting because it implies that the ideal rate of evasion is decreasing in income. The analysis can be extended to this case, but it seems less interesting from an empirical point of view.

[53] A group may be defined by more than one characteristic, corresponding for instance to a specific business sector, the age of the business and a specific region.

chooses a level of evasion $e_i$ for a customer $i$ from class $k$ is:

$$U_j(F, e_i, d_j) = F_j(e_i Y_i) - C(e_i, Y_i, d_j) \tag{3}$$

where $C(e_{i,k}, Y_i, d_j)$ is again the "ethical" cost of allowing $e_i$ evasion for a customer of category $k$ with income $Y_i$, $d_j$ is the accountant's type, and as before $C(e_i, Y_i, m_j) = \frac{(e_i)^2}{2d_i}(Y_i)^\eta$. Accountants are ordered according to their disposition to violate the law with $d_l > d_k$ if $l > k$, $d_j \in [\underline{d}, \overline{d}]$. With a slight abuse of notation, we denote the set of accountants as $J$. Taxpayers and accountants share the net expected monetary benefit of the tax evasion, with the accountant receiving a fraction $\alpha$ of it, implying $F_j(Y_i e_i) = \alpha(1 - (1+T)p_j) Y_i e_i$, which can be positive or negative.

We assume that neither the tax authority nor the TA can observe the taxpayers' types $m_i$. The tax authority however may observe the accountant type $d_j$ and the group $k \in K$ of the taxpayer.[54] We assume the TA targets its auditing effort along these two dimensions. The tax authority chooses the auditing rate for type $k$ of taxpayers served by TA $j$ to maximize expected revenue collection net of the cost of the auditing. If $z_{j,k}$ dollars are spent in auditing a taxpayer of class $k$ assisted by accountant $j$, the share of audited taxpayers is given by the production function $p(z_{j,k}) = \sqrt{z_{j,k}}$. The expected benefit for the tax authority is $p(z_{j,k}) \cdot ((1+T) \cdot E_{j,k} + \xi_{j,k})$, where $E_{j,k} = E(e_i Y_i; j, k)$ is the expected evasion of a taxpayer $i$ of observable type $k$ in TA $j$. The variable $\xi_{j,k}$ is an i.i.d. realization reflecting idiosyncratic factors concerning accountant $j$ and class $k$ that may affect the tax authority's decision.[55] We assume the distribution of $\xi_{j,k}$ is a truncated normal that takes only non-negative values, with mean $\overline{\xi} > 0$ and variance $1/r$. For simplicity, we assume $r$ is sufficiently large that for all practical purposes $\xi_{j,k}$ can be assumed to be normal with $\overline{\xi}_{j,k} > 0$ and variance $1/r$, which allows us to simplify the analysis. The cost of the audit is $\lambda z_{j,k}$, where $\lambda$ is the shadow cost of public funds. Naturally it must be that $z_{j,k} \leq 1$ for all $j$ (or else the probabilities of a discovery will be higher than one). In the following, we assume that, as is natural, $\lambda$ is sufficiently large such that this is always true. For simplicity, we will therefore ignore the constraint $z_{j,k} \leq 1$ going forward.

The timing of the game is as follows. In the first stage the tax authority chooses the auditing strategy $z_{j,k}$, contingent on the observable class of the taxpayer and the TA. In the second stage, the taxpayers choose a TA without observing the tax authority's auditing strategy. In the third stage, each TA $j$ observes $L_k$ informative signals $\mathbf{s_{j,k}} = (s_{j,k,l})_{l=1}^{L}$ on the

[54] Intuitively, the tax authority can observe the TA's activity with many clients over time, thus it can collect more accurate information on the TA's type.

[55] Many factors affect the decision to audit a taxpayer, including the business cycle and the sector in which the taxpayer operates. To these factors, we can add other unobserved factors such as the availability of tax inspectors and general guidelines periodically sent by the Treasury.

auditing strategy $z_{j,k}$ for each class $k$ of taxpayers, and chooses the level of tax evasion for observable class of customer. We assume that each signal $s_{j,k,l} = p^*_{j,k} + \varepsilon_l$, $l = 1, ..., L$ where $\varepsilon_l$ is an i.i.d. normal random variable with mean zero and variance $1/\nu$, and $p^*_{j,k} = p(z^*_{j,k})$ is the actual audit probability. The idea is that the TA can infer this probability by observing a small sample from his/her audited clients. For simplicity we assume $L_k = L$ for all $k$.

We study the perfect Bayesian equilibria in pure strategies of this game. A strategy for a taxpayer is a function $\varphi(\theta_i)$ mapping the taxpayer's type $\theta_i = (m_i, Y_i, k)$ to a TA $j$. A strategy for the tax authority is an allocation of available resources $\mathbf{z} = \left( (z_{j,k})_{j=1}^{J} \right)_{k=1}^{K}$ such that $z_{j,k} \in [0, 1]$ given the observed vector of shocks $\xi_{j,k}$.[56] A strategy for a TA is given by a pair of functions $e(k; d_j, \theta_i, \mathbf{s}_j)$ and $\mu_j(k; \mathbf{s}_{j,k})$. The function $e(k; d_j, \theta_i, \mathbf{s}_{j,k})$ maps the accountant's type, the customer's type and the observed vector of signals $\mathbf{s}_j = (s_{j,1}, ..., s_{j,L})$ to a share of tax evasion in $[0, 1]$ for taxpayer with type $\theta_i$. The function $\mu_j(k; \mathbf{s}_{j,k})$ maps the observed vector of signals to a posterior distribution on the level of auditing chosen by the tax authority for class $k$. This belief is part of the equilibrium because it depends on the TAs beliefs on the tax authority's auditing strategy, given the observed signals.

**Equilibrium behavior.** We solve the game by backward induction. In the last stage, the accountant of type $d_j$ chooses $e_{j,k}$ to maximize (3). From the first order condition we obtain:

$$e(k; d_j, \theta_i, \mathbf{s}_{j,k}) = \begin{cases} \alpha d_j \begin{pmatrix} 1 \\ -(1+T)\, E\left[p(z_{j,k}, \xi_{j,k}); \mathbf{s}_{j,k}\right] \end{pmatrix} (Y_i)^{1-\eta} & \text{if } \begin{array}{c} E\left[p(z_{j,k}, \xi_{j,k}); \mathbf{s}_{j,k}\right] \\ < 1/(1+T) \end{array} \\ 0 & else \end{cases} \tag{4}$$

where $E\left[p(z_{j,k}, \xi_{j,k}); \mathbf{s}_{j,k}\right]$ is the expected level of auditing for class $k$ given the signals $\mathbf{s}_{j,k} = \{s_{j,k,1}, ..., s_{j,k,L}\}$. The accountant chooses a positive level of tax evasion only if the expected probability of auditing or the penalty $T$ are sufficiently small. In this case, the level of tax evasion is decreasing in $E\left[p(z_{j,k}, \xi_{j,k}); \mathbf{s}_{j,k}\right]$. The level of evasion $e(k; d_j, \theta_i, \mathbf{s}_{j,k})$ depends on the accountant's type, and the taxpayer's type, that includes the observable class $k$ and income $Y_i$.

To find the equilibrium, we follow a guess-and-verify approach where we first assume that the equilibrium belief $\mu_j(k; \mathbf{s}_{j,k})$ is such that $p(z_{j,k}, \xi_{j,k})$ follows a normal distribution with mean equal to $A_{j,k}$ and precision $B_{j,k}$. We will then verify that this expectation is correct in equilibrium.

When the accountant believes that $p(z_{j,k}, \xi_{j,k})$ is a $N(A_{j,k}, B_{j,k})$ random variable, Bayes'

[56] Implicitly, the tax authority has a budget R such that $\sum z_j = R$. This is captured by the fact that the cost of choosing $z_j$ is $\lambda$, a parameter that can be interpreted as the Lagrangian multiplier associated with the budget.

rule implies that the posterior probability of the probability of auditing, conditional on the sample $\mathbf{s}_j$, is normally distributed with mean and variance:

$$E\left[p(z_{j,k},\xi_{j,k});\mathbf{s}_{j,k}\right] = \Phi(\overline{s}_{j,k,L}) = \frac{A_{j,k}B_{j,k}+L\nu\cdot\overline{s}_{j,k}}{B_{j,k}+L\nu} \tag{5}$$

$$Var\left[p(z_{j,k},\xi_{j,k});\mathbf{s}_{j,k}\right] = \left(B_{j,k}^{-1}+L\nu\right)^{-1} \tag{6}$$

where $\overline{s}_{j,k}$ is the sample mean of the $L$ signals.[57] Intuitively, the posterior belief is an average of the equilibrium belief on the strategy followed by the tax authority and the evidence collected in the field, i.e. the average signals $\overline{s}_{j,k}$.

In the first stage, the tax authority chooses the amount to spend on auditing $j$'s customers, $z_{j,k}$, given the equilibrium strategy and beliefs of the TAs as described by (4) and (5). The tax authority's problem can be directly written as:

$$\max_{\mathbf{z}\geq 0}\sum\nolimits_j\sum\nolimits_k E\left\{\sqrt{z_{j,k}}\left[(1+T)\left[\begin{array}{c}\alpha d_j\left(1-(1+T)\,\Phi_j(\overline{s}_{j,k})\right)\\ \cdot E\left((Y)^{2-\eta};k\right)\end{array}\right]+\xi_{j,k}\right]-\lambda z_{j,k}\right\} \tag{7}$$

where the expectation reflects the fact that the tax authority does not know the actual sample of signals $\overline{s}_{j,L}$ observed by the consultant; and $E\left((Y)^{2-\eta};k\right)$ is the conditional expectation of $Y^{2-\eta}$. Note that we have:

$$E\Phi(\overline{s}_{j,L}) = \frac{A_{j,k}B_{j,k}+L\nu\cdot(E\overline{s}_{j,k})}{B_{j,k}+L\nu} = \frac{A_{j,k}B_{j,k}+L\nu\cdot\left(\sqrt{z_{j,k}}\right)}{B_{j,k}+L\nu} \tag{8}$$

Substituting (8) in (7), the authority's problem can be directly written in terms of the auditing probabilities $\mathbf{p}=(p_j)_{j=1}^J$:

$$\max_{\mathbf{p}\geq\mathbf{0}}\sum\nolimits_j\sum\nolimits_k E\left[\begin{array}{c}p_{j,k}\left[(1+T)\left[\begin{array}{c}\alpha d_j\left(1-(1+T)\,\frac{A_{j,k}B_{j,k}}{B_{j,k}+L\nu}\right)\\ \cdot E\left((Y)^{2-\eta};k\right)\end{array}\right]+\xi_{j,k}\right]\\ -\left(\frac{\alpha d_j(1+T)^2L\nu}{B_{j,k}+L\nu}E\left((Y)^{2-\eta};k\right)+\lambda\right)p_j^2\end{array}\right]$$

In choosing the optimal $\mathbf{p}$, the tax authority internalizes that there is an additional cost besides $\lambda$: the fact that the TAs change their strategy reducing evasion for the groups that are targeted, the term $\frac{\alpha d_j(1+T)^2L\nu}{B_{j,k}+L\nu}E\left((Y)^{2-\eta};k\right)$ in the second line of the previous expression. Because we assumed above that $\xi_{j,k}$ is positive with arbitrarily high probability, we will

[57]See, for instance Theorem 1 in DeGroot, 1970[ch. 9.5].

ignore for now the cases in which $\xi_{j,k} < 0$, and so:

$$p(z^*_{j,k}, \xi_{j,k}) = \frac{\left[\begin{array}{c} \alpha d_j(1+T)\left(1-(1+T)\frac{A_{j,k}B_{j,k}}{B_{j,k}+L\nu}\right) \\ \cdot E\left((Y)^{2-\eta};k\right) \end{array}\right]}{2\left(\frac{\alpha d_j(1+T)^2 L\nu}{B_{j,k}+L\nu}E\left((Y)^{2-\eta};k\right)+\lambda\right)} + \frac{\xi_{j,k}}{2\left(\frac{\alpha d_j(1+T)^2 L\nu}{B_{j,k}+L\nu}E\left((Y)^{2-\eta};k\right)+\lambda\right)} \tag{9}$$

From the previous expression, the auditing probability $p(z^*_{j,k}, \xi_{j,k})$ can be approximated by a normal random variable with mean equal to the expected value of the right-hand side of (9) and variance equal to the variance of the second term in the right-hand side of (9). In equilibrium, we need that the TAs' beliefs are correct, implying:

$$\begin{aligned} A_{j,k} &= \frac{\alpha d_j(1+T)\left(1-(1+T)\frac{A_{j,k}B_{j,k}}{B_{j,k}+L\nu}\right)E\left((Y)^{2-\eta};k\right)}{2\left(\frac{\alpha d_j(1+T)^2 L\nu}{B_{j,k}+L\nu}E\left((Y)^{2-\eta};k\right)+\lambda\right)} + \frac{\bar{\xi}}{2\left(\frac{\alpha d_j(1+T)^2 L\nu}{B_{j,k}+L\nu}E\left((Y)^{2-\eta};k\right)+\lambda\right)} \\ B_{j,k} &= \left[\left(2\left(\frac{\alpha d_j\left(1+T\right)^2 L\nu}{B_{j,k}+L\nu}E\left((Y)^{2-\eta};k\right)+\lambda\right)\right)^2 \cdot r - L\nu\right]^{-1} \end{aligned} \tag{10}$$

for all $k$ and accountants$j$ who allow their customers of group $k$ to have positive tax evasion.

It is natural that if $\bar{\xi}$ and/or $T$ are large, and/or $\lambda$ small, then the cost of evasion is high and tax evasion may be zero for some $k$. In general, however, it is realistic to assume that the rate of auditing is small since $\lambda$ is large. In these cases we do not need to bother about zero evasion: there is an equilibrium in which the probability of auditing will be sufficiently small so that $e(k; d_j, \theta_i, \mathbf{s}_{j,k}) > 0$ and the equilibrium is characterized by (10).

**Proposition 1.** *There is a $\lambda^*$ such that for $\lambda \geq \lambda^*$ the equilibrium is characterized by (10): the tax authority monitors group $k$ of accountant $j$ with probability $p(z_{j,k}, \xi_{j,k})$ defined in (9), with mean $A_{j,k}$ and variance $B_{j,k}$ solving the system of equations (10); TA $j$ chooses a level of evasion $e(k; d_j, \theta_i, \mathbf{s}_{j,k})$ given by (4) and (5).*

While the equilibrium level of evasion is not, in general, expressible in closed form, it is easy to characterize it in the limit case in which the accountant observes a large number of signals, or very precise signals. In the limit as $L \to \infty$, the expected rate of evasion $e(k; d_j, \theta_i, \mathbf{s}_{j,k})$ converges to $e(k; d_j, \theta_i)$ with:

$$e(k; d_j, \theta_i) = \begin{cases} \Gamma^{\infty}_{j,k}(\xi_{j,k}) \cdot (Y_i)^{1-\eta} & \text{if } \Gamma_{j,k} > 0. \\ 0 & else \end{cases} \tag{11}$$

where

$$\Gamma_{j,k}^{\infty}\left(\xi_{i,k}\right)=\alpha d_{j}\left(1-(1+T)\frac{\left[\alpha d_{j}(1+T)E\left(\left(Y\right)^{2-\eta};k\right)+\xi_{j,k}\right]}{2\left[\alpha d_{j}\left(1+T\right)^{2}E\left(\left(Y\right)^{2-\eta};k\right)+\lambda\right]}\right)$$

Note that as $L\to\infty$, we have $\overline{s}_{j,k}\overset{p}{\to}\sqrt{z_{j,k}^{*}}$ by the law of large numbers, so it is as if the TA could see $z_{j,k}^{*}$. The rate of evasion is a function of the observable demographic class $k\in K$ to which the tax belongs, the ethical standard $d_i$ of the TA, and on the shock $\xi_{j,k}$. The tax authority is aware that TAs with different levels of $d_i$ allow their clients to evade different amounts, but when $\lambda$ is sufficiently high, it is unable (given the cost of resources $\lambda$) to eliminate the heterogeneity in behavior. Note that when, as realistic, $\lambda$ is high, then $\Gamma_{j,k}^{\infty}\left(\xi_{i,k}\right)$ is strictly increasing in $d_j$: so *ceteris paribus* TAs with higher $d_j$ have higher evasion.

Given (11), when $\lambda$ is sufficiently large (so $\Gamma_{j,k}^{\infty}\left(\xi_{i,k}\right)>0$), total evasion is $E_{j,k}=\Gamma_{j,k}^{\infty}\left(\xi_{i,k}\right)\cdot\left(Y_i\right)^{2-\eta}$. The probability distribution of evasion of a category $k$ taxpayer in TA $j$ therefore is:

$$\begin{aligned}\Phi_{j,k}(E)=Pr\left(E_{i,k}\le E\right)=Pr\left(Y_i\le\left(\frac{E}{\Gamma_{j,k}^{\infty}}\right)^{\frac{1}{2-\eta}}\right)&=1-\left(\frac{\left(\underline{Y}^{k}\right)^{2-\eta}\cdot\Gamma_{j,k}^{\infty}\left(\xi_{i,k}\right)}{E}\right)^{\frac{\rho_k}{2-\eta}}\\&=P(E;\underline{E}_{j,k},\widehat{\rho}_k)\end{aligned}$$

where $\underline{E}_{j,k}=\left(\underline{Y}^{k}\right)^{2-\eta}\cdot\Gamma_{j,k}^{\infty}\left(\xi_{i,k}\right)$,$\widehat{\rho}_k=\frac{\rho_k}{2-\eta}$,and $P(E;\underline{E},\widehat{\rho})$ is the Pareto distribution with minimum $\underline{E}$ and shape parameter $\widehat{\rho}$. It follows that $E_{j,k}$ is distributed as a Pareto with minimum value $\left(\underline{Y}^{k}\right)^{2-\eta}\cdot\Gamma_{j,k}^{\infty}\left(\xi_{i,k}\right)$ and shape parameter $\frac{\rho_k}{2-\eta}$. When $\lambda$ is sufficiently large, $\left(\underline{Y}^{k}\right)^{2-\eta}\cdot\Gamma_{j,k}^{\infty}\left(\xi_{i,k}\right)$ is increasing in $j$ and $k$[58] This implies that not only the expected evasion of a taxpayer with a higher observable class $k$ and/or TA with a higher index $d_j$ is higher, but that the entire distribution is shifted to the right in the sense of the first order stochastic dominance.

To make predictions on the average evasion in a TA, we need to characterize how the agents sort between TAs. We have:

**Proposition 2.** *There is a $\lambda^*$ such that for $\lambda\ge\lambda^*$ the equilibrium is characterized by a partition of taxpayers types $\{\widehat{m}_{j,k}\}_{j=1}^{J}$* with $\widehat{m}_{1,k}=\underline{m}$, $\widehat{m}_{J,k}=\overline{m}$ and $\widehat{m}_{j,k}<\widehat{m}_{j+1,k}$ *such that a taxpayer with income $Y_i$, observed class $k$, and ethical standard $m_i$ selects TA $j$ and evades and expected share $e(k;d_j,\theta_i,\mathbf{s}_{j,k})$ as defined in (4),(5) and (10) if $m_i\in(\widehat{m}_{j,k},\widehat{m}_{j+1,k}]$.*

Proposition 2 shows that conditioning on the observable type $k$, taxpayers with a higher propensity to evade (higher $m_i$) match with accountants that are more likely to allow them to do it (higher $d_j$), which causes the distribution of tax evasion to be systematically de-

[58]As $\underline{Y}^{k}$ increases, the tax authority increases the audits, so $\Gamma_{j,,k}^{\infty}$ decreases. When $\lambda$ is large, the effect is small and it is more than compensated by the increase in $\underline{Y}^{k}$.

pendent on the identity of the accountant. Figure 9 illustrates the equilibrium described in Proposition 2.

Given $e(k; d_j, \theta_i)$, in (11), we can write the formulas for the average evasion rate of the customers of TA's $j$:

$$AE(d_j) = \sum_{k \in K} \Gamma^{\infty}_{j,k}\left(\xi_{i,k}\right) E\left(Y^{1-\eta}; k\right) \varphi_{k,j}, \tag{12}$$

where $\varphi_{k,j}$ is the share of $j$'s clients in class $k$ (that will be characterized in the next result). Naturally $AE_j$ is observable only if all customers of a TA are audited, an unlikely event. More useful for our empirical analysis is the average observed evasion of TA $j$, $AOE_j$, defined as:

$$AOE(d_j) = \sum_{k \in K} \Gamma^{\infty}_{j,k}\left(\xi_{i,k}\right) E\left(Y^{1-\eta}; k\right) \cdot E\left(\frac{p(z^*_{j,k}, \xi_{j,k}) \cdot \varphi_{k,j}}{\sum_k p(z^*_{j,k}, \xi_{j,k}) \cdot \varphi_{k,j}}\right) \tag{13}$$

where the expectation in $E\left(\frac{p(z^*_{j,k}, \xi_{j,k}) \cdot \varphi_{k,j}}{\sum_{k'} p(z^*_{j,k'}, \xi_{j,k'}) \cdot \varphi_{k',j}}\right)$ is taken with respect to the realization of the vector $\xi_j = \left(\xi_{j,k}\right)_k$.

In general the size of the brackets $\Delta^m_{j,k} = \widehat{m}_{j+1,k} - \widehat{m}_{j,k}$ in which the taxpayers select accountant $j$ may depend on the observable type $k$ of the taxpayer. The brackets $\Delta^m_{j,k}$ may differ in size as we change $k$ because different observable types may expect different audit rates, which may affect their preference for the type of TA. When, as it happens in practice, $\lambda$ is large and so the probability of an audit is small, these differences are small and $\Delta^m_{j,k}$ changes little in $k$. In these cases, the distribution of observable types is relatively constant across TAs of different types. Note that:

$$z_{j,k} = E\left(\frac{p(z^*_{j,k}, \xi_{j,k}) \cdot \varphi_{k,j}}{\sum_k p(z^*_{j,k}, \xi_{j,k}) \cdot \varphi_{k,j}}\right)$$

is a probability distribution function since $z_{j,k} \in [0,1]$ and $\sum_k z_{j,k} = 1$. We say that $Z_j$ first order stochastically dominates $Z_{j'}$ if for every $l$, $Z_j(l) = \sum_{k \geq l} z_{j,k} \geq Z_{j'}(l)$. In this case we write $Z_j \succeq Z_{j'}$. We have:

**Proposition 3.** *There is a $\lambda^*$ such that for $\lambda \geq \lambda^*$, we have $Z_j \succeq Z_{j'}$ when $j \geq j'$, $AOE(d_j) > AE(d_j)$, and $AE(d_j)$ and $AOE(d_j)$ are both strictly increasing in $d_j$.*

Proposition 3 implies that we should expect the evasion of taxpayer $i$ to be positively correlated with both the average evasion of the other customers of $i$'s TA and in the average evasion of the other customers of $i$'s accountant that are audited in equilibrium.

Summarizing, Propositions 1-3 lead to the following observations. First, we have:

**Observation 1. (Sorting Effect)** *Except when the cost of auditing is zero, taxpayers with* *a higher propensity to evade match with TAs who are more accommodating in equilibrium. This implies that the expected rate of tax evasion of a taxpayer is increasing in both the rate of evasion of other customers of the taxpayers' accountant, and in the observed rate of evasion of other customers of the taxpayer's accountant who are audited in equilibrium.*

A second important implication of the model is that the final level of tax evasion depends not only on the accountant's type, but also on the information that the accountant acquires regarding the auditing strategy followed by the tax authority. The TA fine-tunes the level of tax evasion based on the accountant's tolerance for evasion (i.e. the type $d_j$) and the observed average signals $\overline{s}_{j,k}$. When $d_j$'s are positive (i.e. when there is some tolerance for evasion), we observe heterogeneity in behavior due to heterogeneous signals: a TA who receive a higher (lower) signal $\overline{s}_{j,k}$ reacts by reducing (increasing) the evasion of all other customers of type $k$ (see (4) and (5)) This leads to correlated behavior among the customers of a TA even if there is no heterogeneity in the types $d_i$s of the accountants. For a given true auditing policy $(z_{j,k})_{j,k}$, a higher (lower) signal does not change the auditing probability, it only changes the rate of evasion: so it increases both the true rate of evasion and the rate of evasion that is observed by the audits. This can be directly seen in the limit case represented by (11): here the TA observes directly the auditing rate. The auditing rate changes in equilibrium if there is a change in $\xi_{i,k}$. A reduction/increase in $\xi_{i,k}$ is reflected in (11) in a reduction/increase in $\Gamma^{\infty}_{j,k}(\xi_{i,k})$ and therefore both in a first order stochastic dominance change in the entire distribution of evasion, and in a change in $AE(d_j)$ and $AOE(d_j)$. We have:

**Observation 2. (Informational Externality Effect I)** *Even if there is no sorting because all TAs have the same type $d_j = d^* > 0$, the expected rate of tax evasion of a taxpayer is increasing in both the rate of evasion of other customers of the taxpayer's accountant, and in the rate of evasion of other customers of the taxpayer's accountant who are audited in equilibrium.*

Without directly observing the TA's types, it would be hard to separate the sorting effect vs. the informational spillover effect and thus test Observations 1 and 2. The informational spillover effect, however, has two additional testable implications. The most likely signal used by the TA to fine-tune his/her activities at time $t$ is his/her direct experience with customers at time $t-1$ and perhaps the experience of nearby accountants if they can communicate. We can therefore interpret the average signal $\overline{s}_{j,k}$ as the number of $j$'s customers of type $k$ who are audited at time $t-1$. Given this, we tax practitioner should expect the expected probability of receiving an audit at time $t$ to be increasing with the number of clients that are audited at $t-1$; or, if accountant $j$ is in communication with an accountant $l$, increasing with the share of $l$'s clients of observable type $k$ who are audited. Hence, It follows that:

**Observation 3. (Informational Externality Effect II)** *The reported income (evasion) of a taxpayer at $t$ is expected to correlate positively (negatively) with the number of other clients of the same tax practitioner audited at $t-1$.*

In the simple model presented above we have assumed that the signals received by TA $j$ with respect to the auditing strategy $z_{j,k}$ are informative only for taxpayers of type $k$ (i.e. signals $s_{j,k,l} = p^*_{j,k} + \varepsilon_l$ for signals $l = 1, ..., L$, where the noise $\varepsilon_l$ is independently distributed across groups $k = 1, ..., K$). In this case a signal $s_{j,k,l}$ reveals information on $z_{j,k}$ but not on $z_{j,k'}$ for $k \neq k'$. It is conceptually straightforward (though more cumbersome) to extend the model to allow the signals to be correlated.[59] In this case we should expect that a higher signal received on group $k$ will also induce less evasion for taxpayers of type $k'$ served by the same tax practitioner. Summarizing:

**Observation 4. (Informational Externality Effect III)** *The behavioral response of taxpayers to audits of other customers of their TA is higher when the other audited taxpayers are similar to them in terms of observables, such as the business sector, size or gender, but it is expected to be present also across classes of taxpayers.*

In practice, we can have four cases. When the heterogeneity in the practitioners types is sufficiently important and the accountants' signals are sufficiently precise, we should observe both self-selection and informational spillovers in the data. We might, however, have three other cases: if practitioners' types $d_j$ are not very heterogeneous, but signals are important, then we might observe only the informational spillover effect; when accountants' types are heterogeneous, but signals are uninformative, then we might observe only the sorting effect; if both types of heterogeneity are weak and signals are uninformative, then we might not observe either of the two effects.

## A.1 Proof of Proposition 1

Assume that all levels of evasion are positive, so that the mean and variance of the accountant's beliefs are a fix point of $(A_{j,k}, B_{j,k}) = F_k(A_{j,k}, B_{j,k})$, where $F_k(A_{j,k}, B_{j,k})$ is defined by (10). Note that $F_k$ is continuous in $A_{j,k}, B_{j,k}$ and $A_{j,k}, B_{j,k}$ must be in $[0, 1] \times [0, \overline{B}]$ where

$$\overline{B} = \left[(2\lambda)^2 \cdot r - L\nu\right]^{-1}$$

[59] A closed form solution for the posterior in the case of correlated signals similar to (3) is presented in DeGroot, 1970, see Theorem 1 in Section 9.9.

It follows by the Kakutani fix point theorem that $A_{j,k}, B_{j,k}$ exists, and $A_{j,k}, B_{j,k}$ are both strictly positive. By construction, the TA choice of tax evasion is optimal given the beliefs and the beliefs are correct in equilibrium. Similarly, the tax authority chooses the optimal level of auditing given the correct beliefs of the accountant's evasion. It is easy to see that there is a $\lambda^* > 0$ such that for $\lambda > \lambda^*$ the associated equilibrium auditing is sufficiently small that $E\left[p(z_{j,k}, \xi_{j,k}); \mathbf{s}_{j,k}\right] < 1/(1+T)$ is satisfied. $\blacksquare$

## A.2 Proof of Proposition 2

The taxpayer $i$ does not know $\overline{s}_{j,k}$ when choosing consultant $j$, so expects an evasion rate:

$$Ee(k; d_j, \theta_i, \mathbf{s}_{j,k}) = \alpha d_j \left(1 - (1+T)\frac{A^*_{j,k}B^*_{j,k} + L\nu \cdot Ez^*_{j,k}}{B^*_{j,k} + L\nu}\right)(Y_i)^{1-\eta} = \Gamma^e_{j,k} \cdot (Y_i)^{1-\eta}$$

where $A^*_{jk}$, $B^*_{j,k}$ and $z^*_{j,k}$ are the equilibrium levels of $A_{j,k}$, $B_{j,k}$ and $z_{j,k}$ given by (10) . Taxpayers choose $j$ to maximize:

$$(1-\alpha)\left(1 - (1+T)\, E\left[p(z_{j,k}, \xi_{j,k})\right]\right)\Gamma^e_{j,k} \cdot (Y_i)^{2-\eta} - \frac{\left(\Gamma^e_{j,k} \cdot (Y_i)^{1-\eta}\right)^2}{2m_i}(Y_i)^{\eta}$$

Note that there is a $\lambda^*$ such that for $\lambda \geq \lambda^*$, $\left(1 - (1+T)\, E\left[p(z_{j,k}, \xi_{j,k}); \mathbf{s}_{j,k}\right]\right)\Gamma^e_{j,k} \cdot (Y_i)^{2-\eta}$ is strictly increasing in $d_j$, since $E\left[p(z_{j,k}, \xi_{j,k}); \mathbf{s}_{j,k}\right]$ and $E\left[p(z_{j,k}, \xi_{j,k})\right]$ become insensitive to $j$. It follows that $(1-\alpha)\left(1 - (1+T)\, E\left[p(z_{j,k}, \xi_{j,k})\right]\right) Y_i e(d_j))$ is increasing in $d_j$ as well for $\lambda \geq \lambda^*$. Assume now that type $m_i$ of observable type $k$ with income $Y_i$ prefers $d_j$ to $d_l$ with $d_j > d_l$. Then we have:

$$m_{\tau,k} \geq m_{i,k} \geq \frac{1}{2}\left(\frac{\left(\Gamma^e_{j,k}\right)^2 - \left(\Gamma^e_{l,k}\right)^2}{\begin{bmatrix} (1-\alpha)\left[\left(1 - (1+T)\, E\left[p(z_{j,k}, \xi_{j,k})\right]\right)\left(\Gamma^e_{j,k}\right)\right] \\ \left(1 - (1+T)\, E\left[p(z_{l,k}, \xi_{l,k})\right]\right)\left(\Gamma^e_{l,k}\right) \end{bmatrix}}\right)$$

for any $m_{\tau,k} \geq m_{i,k}$, implying that a taxpayer of type $m_{\tau,k}$ with income $Y_\tau$ also prefers $d_j$ to $d_l$. Similarly, we can show that if $m_i$ of type $k$ prefers $d_j$ to $d_k$ with $d_j < d_k$, then $m_l$ of type $k$ also prefers $d_j$ to $d_k$ for any $m_l \leq m_i$. This implies that the set of $m_i$s who chooses a consultant $d_j$ is convex, increasing in $d_j$ and independent of income $Y_i$. That is there is a set of cut points $\{\widehat{m}_{j,k}\}_{j=1}^{J}$ with $\widehat{m}_{j,k} \leq \widehat{m}_{j+1,k}$ and $\widehat{m}_{j,k} \in [\underline{m}, \overline{m}]$, such that all types in $(\widehat{m}_{j,k}, \widehat{m}_{j+1,k}]$ find it optimal to choose a consultant of type $d_j$. $\blacksquare$

## A.3 Proof of Proposition 3

From (10) we can see that both $A_{j,k}$ and $B_{j,k}$ converge to zero as $\lambda \to \infty$, so by (4), so $E\left[p(z_{j,k},\xi_{j,k});\mathbf{s}_{j,k}\right] \to 0$ and $\Gamma_{j,k}{\to}\alpha d_j$ as $\lambda \to \infty$. This implies that:

$$\hat{m}_{j,k} \to \frac{1}{2}\left(\frac{\alpha}{(1-\alpha)}\right)\left[d_j + d_{j+1}\right]$$

and:

$$\hat{m}_{j+1,k} - \hat{m}_{j,k} \to \frac{1}{2}\left(\frac{\alpha}{(1-\alpha)}\right)\left[d_{j+2} - d_j\right] = \hat{m}_{j+1} - \hat{m}_j$$

which are both independent of $k$. Moreover from (12):

$$AE(d_j) \to \sum_{k\in K}\left[\alpha d_j\right]\cdot E\left(Y^{1-\eta};k\right)\frac{(\hat{m}_{j+1}-\hat{m}_j)}{K}$$

which is strictly increasing in $j$. We conclude that there is a $\lambda^*$ such that $AE(d_j)$ is strictly increasing in $j$ for $\lambda > \lambda^*$.

Consider now $AOE(d_j)$, we have:

$$\begin{gathered} E\left[\frac{p(z^*_{j,k},\xi_{j,k})}{\sum_k p(z^*_{j,k},\xi_{j,k})}\right] \to E\left[\frac{\alpha d_j(1+T)E\left((Y)^{2-\eta};k\right)+\xi_{j,k}}{\sum_{k''}\left[\alpha d_j(1+T)E\left((Y)^{2-\eta};k''\right)+\xi_{j,k''}\right]}\right] \\ = \int\cdots\int_{\xi_l}\left[\frac{\alpha d_j(1+T)E\left((Y)^{2-\eta};k\right)+\xi}{\sum_{k''}\left[\alpha d_j(1+T)E\left((Y)^{2-\eta};k''\right)+\xi_{k''}\right]}\right] dT(\xi_1)...dT(\xi_{k-1})dT(\xi_{k+1})...dT(\xi_K)\cdot dT(\xi) \end{gathered}$$

where $T$ is the cumulative distribution of $\xi_k$. The last line of the preceding expression is strictly increasing in $k$. Note that for $k > k'$:

$$\begin{gathered} lim_{\lambda\to\infty}E\left[\frac{p(z^*_{j,k},\xi_{j,k})}{\sum_{k''}p(z^*_{j,k''},\xi_{j,k''})}\right] - lim_{\lambda\to\infty}E\left[\frac{p(z^*_{j,k'},\xi_{j,k'})}{\sum_{k''}p(z^*_{j,k''},\xi_{j,k''})}\right] \\ = \left[\alpha d_j(1+T)\left[E\left((Y)^{2-\eta};k\right) - E\left((Y)^{2-\eta};k'\right)\right]\right]\cdot E\left[\frac{1}{\sum_{k''}\left[\alpha d_j(1+T)E\left((Y)^{2-\eta};k''\right)+\xi_{k''}\right]}\right] \\ = \left[\alpha(1+T)\left[E\left((Y)^{2-\eta};k\right) - E\left((Y)^{2-\eta};k'\right)\right]\right]\cdot E\left[\frac{1}{\sum_{k''}\left[\alpha(1+T)E\left((Y)^{2-\eta};k''\right)+\xi_{k''}/d_j\right]}\right] > 0 \end{gathered}$$

and increasing in $d_j$. We conclude that if we increase $d_j$ then we have a first order stochastic dominance improvement in the distribution $z^\infty_{j,k} = lim_{\lambda\to\infty}E\frac{p(z^*_{j,k},\xi_{j,k})}{\sum_k p(z^*_{j,k},\xi_{j,k})}$. Since $E\left(Y^{1-\eta};k\right)$ is increasing in $k$, we can write:

$$\begin{gathered} lim_{\lambda\to\infty}AOE(d_j) = \left[\alpha d_j\right]\sum_{k\in K}E\left(Y^{1-\eta};k\right)\cdot z^\infty_{j,k} \\ > \left[\alpha d_l\right]\sum_{k\in K}E\left(Y^{1-\eta};k\right)\cdot z^\infty_{l,k} = lim_{\lambda\to\infty}AOE(d_l) \end{gathered}$$

for $j > l$. Again we conclude that there is a $\lambda^*$ such that $AOE(d_j)> AOE(d_l)$ for $\lambda > \lambda^*$ for $j > l$. Similarly, since $z^{\infty}_{l,k}$ and $E\left(Y^{1-\eta}; k\right)$ are increasing in $k$, we have $\sum_{k\in K} E\left(Y^{1-\eta}; k\right) \cdot z^{\infty}_{j,k} > \sum_{k\in K} E\left(Y^{1-\eta}; k\right) \cdot (1/K)$, which implies that $AOE(d_j) > AE(d_j)$. ■

# B Information sharing among tax accountants

A TA can update his/her understanding of the audit policy by gathering the signals through the audits on his/her clients, or also by sharing this information with other TAs. Sharing information about IRA audits to his/her clients brings benefits as a TA can offer more precise advice to their own customers, but it comes at the cost of losing competitive advantage as information is passed over to other TAs. Because tax evasion is illegal and TAs are heterogenous in evasion proneness, law abiding TAs may be reluctant to trade information with less lawful colleagues, undermining the benefits of information sharing. Here we provide some evidence of whether TAs share information on their clients audits.[60] If they do, they may be more likely to do so with similar accountants in the same area. To this end, we build on our test in Table 9, and check whether the income of a taxpayer responds to audits on taxpayers who are advised by accountants similar to the own accountant. The results are reported in Table A1. In the first column, we consider the impact of an audit on a taxpayer advised by a TA in the same cluster of the own TA. Clusters are defined using the machine learning algorithm described in Section 4 (Figure 5). We find that there is no spillover effect. This result remains unchanged even if we consider audits on similar non-peers (column 2). Specifically, we find that a taxpayer does not change the income declared after an audit on a similar taxpayer in his/her same cluster, if they are advised by a TA in the same cluster of his/her own TA. These findings suggest that TAs do not share information.[61]

[60]Notice that finding no evidence that TAs do not share information on their clients audits does not mean that TAs do not share information in general; clearly they exchange information, for instance on how to interpret the tax code through their professional association.

[61]We have also run placebo tests when defining TA similarity along specific TA characteristics such as sector and number of clients, as in Figure S1. Specifically, we have estimated our regression model 1,000 times after having randomly assigned each taxpayer to a different accountant in the same province *a*) with at least one client in the same 2-digit sector; or *b*) in the same decile of the size distribution. We have also considered a third similarity dimension: *c*) we have assigned taxpayers to a random TA in the same province and decile of evaders over audited clients. Figure S2 in the Supplementary Appendix shows that the distribution of estimated placebo coefficients is centered around zero. In the vast majority of the cases (98.1% in case *a*, 99.1% in case *b* and 98.6% in case *c*), we find no effect on reported income at $t$ of the share of audited customers of a different accountant at $t$-1.

# C Quantification of Deterrence Effects: Details

In this section, we explain how we compute the total cumulative direct and indirect deterrence effects reported in Section 6. Let $\alpha_i$ be the marginal direct effect of own audit at lag $i = 1, 2, 3$; and let $\beta_i$ be the marginal effect of the others' audits at lag $i = 1, 2, 3$ as estimated in Table 10, column 3. The direct effect is estimated as *0.9*$\times$*Number of audits*$\times(\sum \alpha_i)\times$*Average income of audited* $= 0.9 \times 377,113 \times (0.118) \times 33,743 = 1,351$ million euros, where the average income figure is that of the audited taxpayers who report a positive income; we use this measure because, given our log specification in Table 10, the marginal effects $\alpha_i$ and $\beta_i$ are not defined for the taxpayers that report zero income. The latter are around 10% of the taxpayers, which explains why we multiply the above expressions by 0.9. The cumulative spillover effect of an additional audit is equal to *0.9*$\times$*Number of affected clients of a TA*$\times(\sum \beta_i)\times$*Average income* $= 0.9 \times 30.13 \times (0.019) \times 21,190 = 10,918$ euros, where the average income is that of the total sample conditional on reporting a positive income. The total spillover is obtained by multiplying this number by the number of TAs affected by audits (65,133). The total indirect effect is 711 million euros. In so far own audits and peers audits also affect the probability of reporting positive income (as it does as discussed in Section 5.1), these calculations are a conservative estimate of the true direct and indirect effects.

Additionally, we can compute how tax audits contribute to tax revenue mechanically because they allow to discover evasion and through deterrence effects. To provide a back of the envelope calculation we focus on the audits conducted in year 2012 (81,797 audits) and compute the contribution to total revenue of each channel. The 2012 audits detected 1,347 million euros of evaded taxes; of this, only 433 million were ultimately recovered. This is the mechanical contribution of the audits. The direct deterrence effect corresponds to the tax on the additional income declared in the following years by the audited taxpayer (from the above formula the extra income is: $0.9 \times 81,797 \times (0.118) \times 33,743 = 293$ million euros). By applying the average tax rate on the average income of the audited (38%), the total direct deterrence effect of the audits amounts to 111 million euros. The indirect deterrence effect from the TA spillovers is equal to the tax on the additional average income declared by peers. Using the previous formula, the additional taxable income declared from the peers is 415 million euros ($0.9 \times 30.13 \times (0.019) \times 21,190 \times 38,031$). Considering a tax rate of 27% (the average for the tax bracket of 15,000-28,000 euros), the total indirect deterrence effect of audits in 2012 amounts to 112 million euros. In other words, the tax revenue yield of an audit is due by 2/3 to the tax evasion assessment and by 1/3 to the deterrence effect, equally split between the direct and the TA spillover channels.

# Figures

(A) Aggregated sectors (2-digit) (B) Sectors (5-digit)

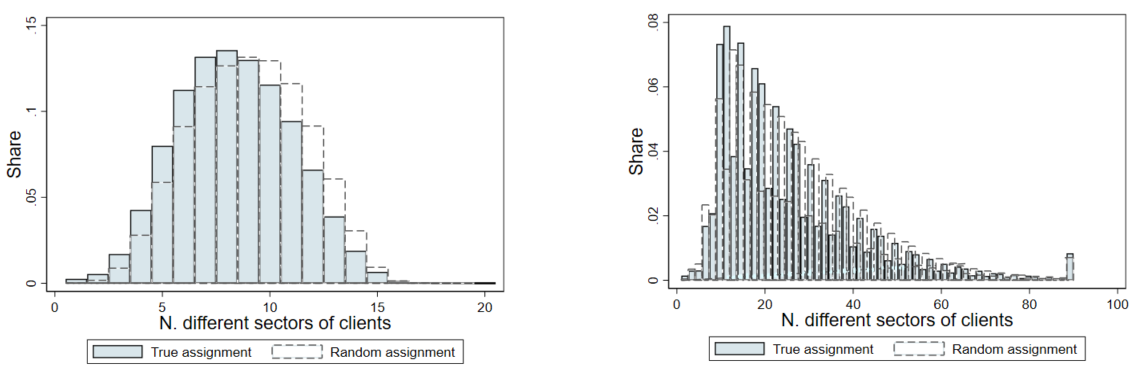


Figure 1
Distribution of TAs by Number of Sectors of Clients

These figures show the number of different sectors of activity of clients of TAs with at least ten clients. Panel A uses the 2-digit classification of sectors, panel B uses the 5-digit classification of sectors. The solid bars report the actual distribution observed in the data, the dashed bars report the distribution obtained by reassigning taxpayers to a random TA in the same province. In panel B, the distribution is winsorized at the 90th percentile.

(A) Income (B) Experience (C) Age

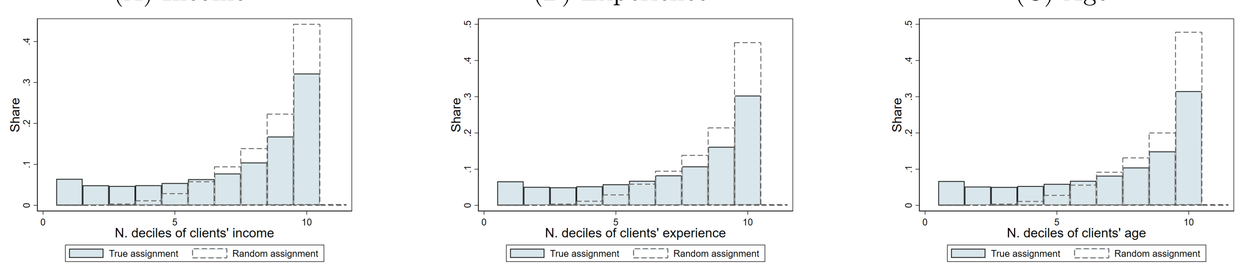


Figure 2
Distribution of TAs by Number of Deciles of Clients' Characteristics

These figures show the number of different deciles of clients' income, experience, and age. In panel A, the income deciles are computed using the distribution of tax filings in the same year, province, and 5-digit sector. The solid bars report the actual distribution observed in the data, the dashed bars report the distribution obtained by reassigning taxpayers to a random TA in the same province.

(A) Share of audited clients

(B) Share of audited clients
with positive evasion

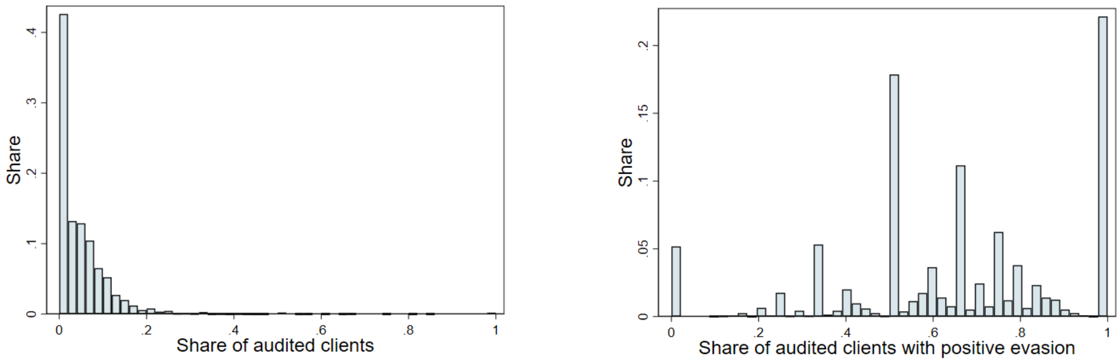


Figure 3
Distribution of TAs by Share of Audited Clients

Panel A shows the share of audited clients of all TAs. Panel B shows the share of audited clients with positive evasion of all TAs with at least two clients audited.

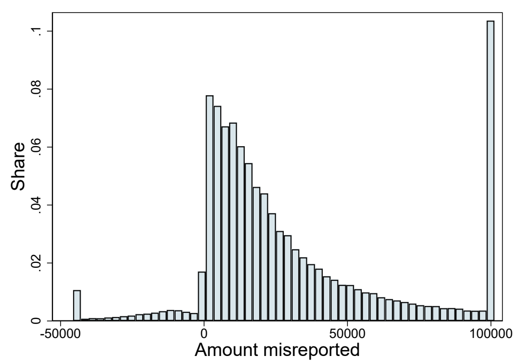


Figure 4
Distribution of Tax Filings by Amount of Misreporting

This figure shows the amount of misreported taxable income winsorized on the left at the 1st percentile and on the right at the 90th percentile. The sample includes all audited tax filings.

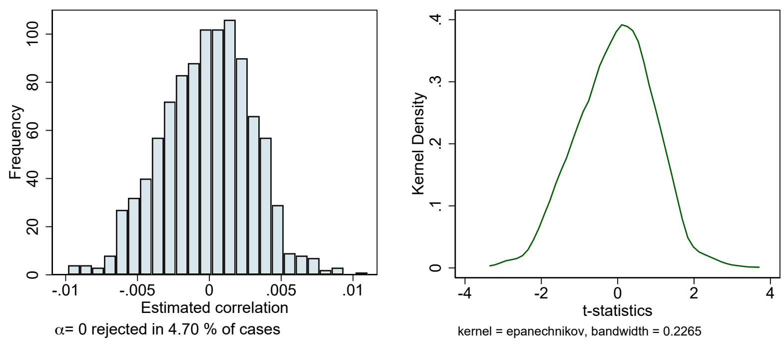


Figure 5
Placebo Regressions - Spillover Effect

This figure shows the distribution of estimated coefficients $\alpha$ and $t$-statistics for the OLS specification in Table 6, column 2, when randomly assigning TAs in the same TA cluster. The spillover estimate obtained in Table 6, column 2, is 0.116.

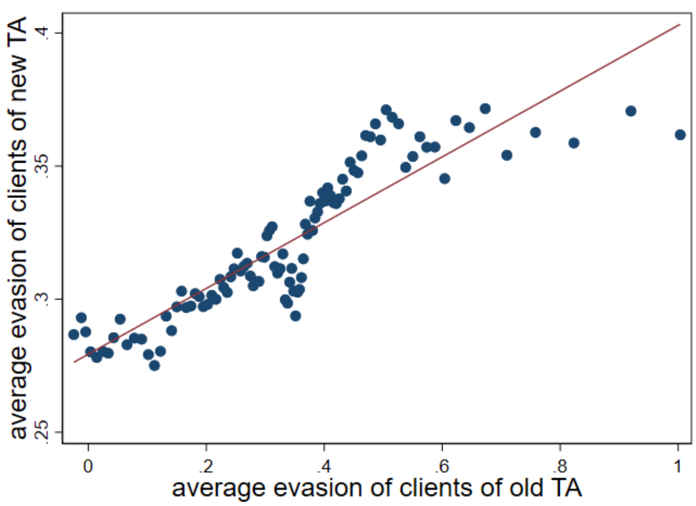


Figure 6
Sorting of Taxpayers into TAs

The $x$-axis shows the mean share of evasion of the clients of the TA of origin binned in percentiles. The $y$-axis reports the mean share of evasion of the clients of the new TA after partialling out the characteristics of the taxpayer and his business, year, and province fixed effects. The sample includes taxpayers changing TA at least once in the observed period.

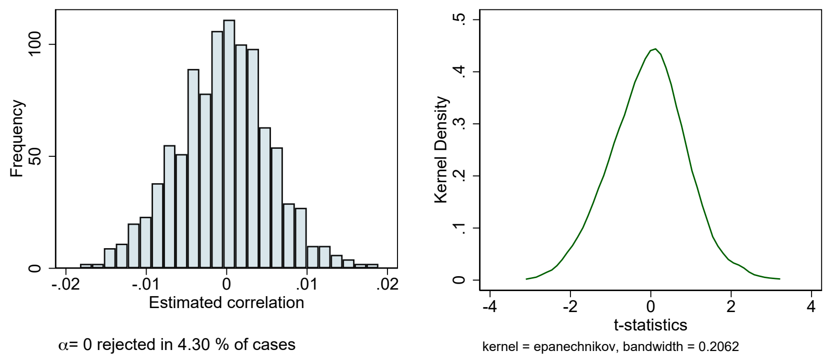


Figure 7
Placebo Regressions - Sorting Effect

This figure shows the distribution of estimated coefficients $\alpha$ and $t$-statistics from the OLS specification in Table 11, column 1, when randomly assigning TAs in the same TA cluster. The estimate of the sorting effect obtained in Table 11, column 1, is 0.042.

(A) Sorting along tax avoidance

(B) Sorting along probability of appeal after an audit

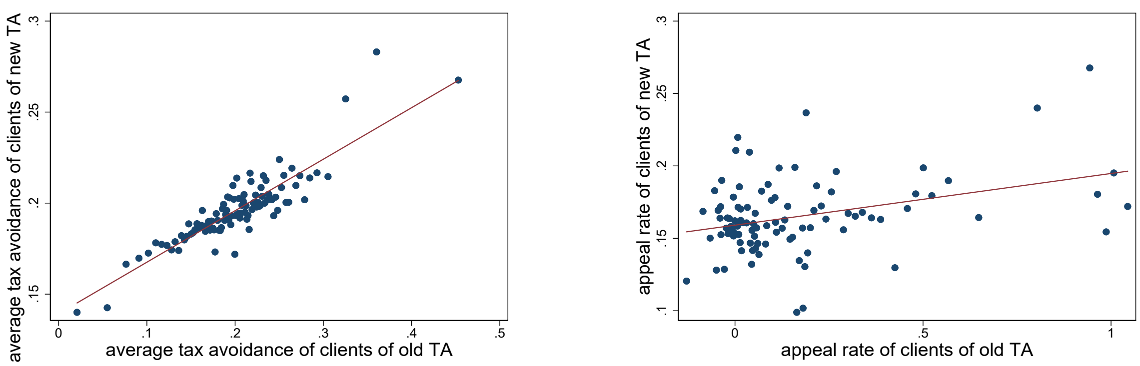


Figure 8
Sorting of Taxpayers into TAs Along Tax Avoidance and Appeal Rate

The x-axis shows the avoidance rate (panel A) or the appeal rate (panel B) of the TA of origin binned in percentiles. The y-axis reports the avoidance rate and the appeal rate of the clients of the new TA, respectively, after partialling out the characteristics of the taxpayer and his business, as well as year, sector and province fixed effects. The sample includes taxpayers changing TA at least once in the observed period. The avoidance rate of a TA is computed as the mean tax avoidance of his/her clients. The appeal rate of a TA is computed as the share of the audited clients with positive detected evasion who appeal the audit assessment.

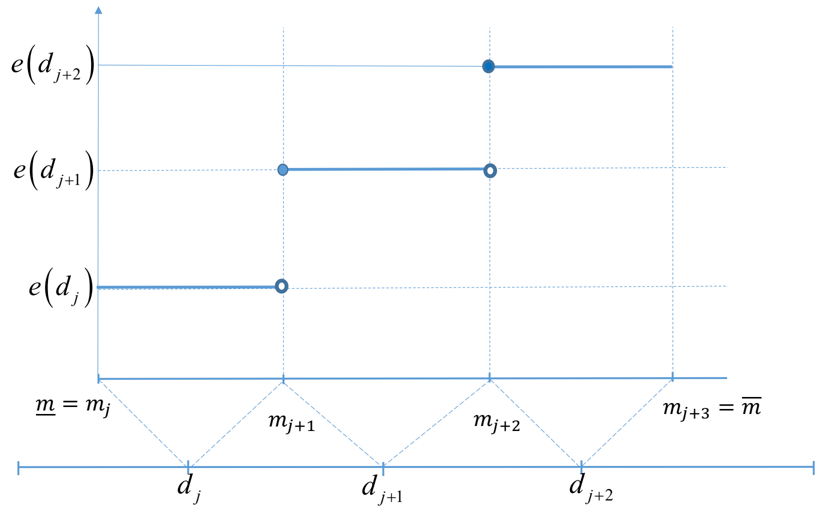


Figure 9
The Equilibrium

# Tables

Table 1
Summary Statistics - Taxpayers and Audits

*A. Taxpayers*

N. taxpayers: 4,697,751

| N. tax filings: 20,324,271 | mean | median | sd | 10th pct | 90th pct |
|---|---|---|---|---|---|
| Woman | 0.27 | 0 | 0.44 | 0 | 1 |
| Married | 0.65 | 1 | 0.48 | 0 | 1 |
| Age | 46.77 | 46 | 12.45 | 32 | 63 |
| Experience (years of activity) | 13.50 | 12 | 10.43 | 1 | 29 |
| Firm size (n. employees) | 0.83 | 0 | 3.19 | 0 | 2 |
| Mover | 0.07 | 0 | 0.25 | 0 | 0 |
| Agriculture | 0.12 | 0 | 0.33 | 0 | 1 |
| Trade | 0.27 | 0 | 0.44 | 0 | 1 |
| Construction and Manufacturing | 0.19 | 0 | 0.39 | 0 | 1 |
| Private services | 0.37 | 0 | 0.48 | 0 | 1 |
| Health, education and recreational services | 0.04 | 0 | 0.20 | 0 | 0 |
| Filed income | 18,640.10 | 10,515 | 48,693.71 | 0 | 39,997 |
| Agriculture | 4,286.36 | 845 | 28,116.77 | 0 | 9,775 |
| Trade | 15,175.54 | 10,002 | 26,548.88 | 0 | 32,342 |
| Construction and Manufacturing | 16,174.32 | 13,164 | 237,854.85 | 246 | 29,856 |
| Private services | 23,913.77 | 12,270 | 69,011.16 | 0 | 49,607 |
| Health, education and recreational services | 47,356.02 | 34,044 | 54,771.84 | 2,819 | 99,356 |
| Tax avoidance | 0.21 | 0.15 | 0.24 | 0 | 0.47 |
| Agriculture | 0.26 | 0.05 | 0.37 | 0 | 1 |
| Trade | 0.23 | 0.20 | 0.23 | 0 | 0.49 |
| Construction and Manufacturing | 0.21 | 0.19 | 0.20 | 0 | 0.39 |
| Private services | 0.19 | 0.13 | 0.22 | 0 | 0.41 |
| Health, education and recreational services | 0.12 | 0.09 | 0.14 | 0.01 | 0.25 |

*B. Audits and Evasion*

N. audited tax filings: 388,513 ; N. audited taxpayers: 289,434
N. audits by *Guardia di Finanza*: 11,400
Taxpayers audited at least once: 6.16%

| Taxpayers with positive evasion: 66.45% | mean | median | sd | 10th pct | 90th pct |
|---|---|---|---|---|---|
| Yearly % not congruent tax filings | 35.06 | 33.22 | 5.40 | 29.47 | 45.17 |
| Yearly % not coherent tax filings | 51.79 | 51.73 | 3.62 | 46.01 | 56.39 |
| Yearly % audited tax filings | 1.89 | 2.42 | 1.28 | 0.19 | 3.35 |
| Agriculture | 0.77 | 0.98 | 0.52 | 0.10 | 1.34 |
| Trade | 2.00 | 2.53 | 1.38 | 0.19 | 3.54 |
| Construction and Manufacturing | 2.02 | 2.34 | 1.51 | 0.13 | 3.89 |
| Private services | 2.08 | 2.82 | 1.39 | 0.22 | 3.75 |
| Health, education and recreational services | 2.16 | 2.58 | 1.24 | 0.42 | 3.51 |
| Age of audited tax filings | 3.87 | 4 | 0.96 | 2 | 5 |
| Audit duration (days) | 110.56 | 60 | 173.26 | 14 | 194 |
| Filed income \| audit | 29,602.44 | 13,560 | 100,753.95 | 0 | 60,327 |
| Evaded income \| audit | 20,328.02 | 4,053 | 143,316.44 | 0 | 36,256 |
| Filed income \| positive evasion | 29,090.98 | 13,700 | 104,127 | 0 | 58,212 |
| Evaded income \| positive evasion | 32,688.89 | 10,139 | 180,624.23 | 2,524 | 56,673 |
| Share of evasion on total income \| audit | 0.33 | 0.20 | 0.35 | 0 | 0.94 |
| Agriculture | 0.30 | 0.09 | 0.36 | 0 | 0.93 |
| Trade | 0.34 | 0.22 | 0.36 | 0 | 0.98 |
| Construction and Manufacturing | 0.38 | 0.30 | 0.36 | 0 | 0.97 |
| Private services | 0.30 | 0.16 | 0.35 | 0 | 0.91 |
| Health, education and recreational services | 0.18 | 0.05 | 0.26 | 0 | 0.60 |
| Share of evasion on total income \| positive evasion | 0.50 | 0.46 | 0.32 | 0.08 | 1 |
| Appeal \| audit | 0.19 | 0 | 0.39 | 0 | 1 |
| Agriculture | 0.23 | 0 | 0.42 | 0 | 1 |
| Trade | 0.18 | 0 | 0.39 | 0 | 1 |
| Construction and Manufacturing | 0.15 | 0 | 0.35 | 0 | 1 |
| Private services | 0.20 | 0 | 0.40 | 0 | 1 |
| Health, education and recreational services | 0.21 | 0 | 0.41 | 0 | 1 |

*Notes.* Income figures are expressed in euros.

Table 2
LASSO Model Selection

| Model | Sample | Non-zero coefficients | Deviance | Deviance Ratio |
|---|---|---|---|---|
| Minimum BIC | Training | 25 | 0.056 | 0.106 |
| | Testing | | 0.054 | 0.099 |
| Cross Validation | Training | 158 | 0.054 | 0.132 |
| | Testing | | 0.053 | 0.107 |
| Adaptive LASSO | Training | 116 | 0.054 | 0.131 |
| | Testing | | 0.053 | 0.107 |

*Notes.* This table reports goodness of fit measures of alternative probit LASSO models of a dummy variable with value one if the tax filing is audited on all available information on tax filings (233 variables). The sample includes all tax filings at risk of audit in any year of audit. Because of the computationally intensive LASSO procedure and the very large size of our sample, the model selection exercise is performed on a 1% random extraction of the sample. Models are trained of a sample of 402,976 observations (50% of the random sample) and then tested out-of-sample on the remaining 402,976 observations. Postselection coefficients are considered. The table reports the number of non-zero coefficients selected by each model and the relative measures of fit (Hastie et al., 2015). The deviance and deviance ratio of the probit model using as covariates the audit policy controls in our baseline estimates (153 variables) in the testing sample are equal to 0.053 and 0.113, respectively.

Table 3
Balance Tests

| | Dep. Var. at $t+1$: | | | | | |
|---|---|---|---|---|---|---|
| ***Panel A: All audits*** | Income | Share of evasion | Tot. taxable revenues | VAT taxable turnover | Operating costs | Operating costs/NPV |
| Audited at $t$ | -705.504** | -0.005* | 2,450.539*** | 1,881.596*** | 2,141.219*** | 1.531*** |
| | (299.423) | (0.003) | (768.985) | (712.426) | (708.736) | (0.331) |
| Audit Policy Controls | no | no | no | no | no | no |
| ***Panel B: All audits*** | Income | Share of evasion | Tot. taxable revenues | VAT taxable turnover | Operating costs | Operating costs/NPV |
| Audited at $t$ | -4.328 | 0.002* | 226.286 | -12.746 | 159.856 | -0.013 |
| | (31.300) | (0.001) | (262.234) | (147.688) | (109.292) | (0.113) |
| Audit Policy Controls | yes | yes | yes | yes | yes | yes |
| ***Panel C: Audits by GdF*** | Income | Share of evasion | Tot. taxable revenues | VAT taxable turnover | Operating costs | Operating costs/NPV |
| Audited at $t$ | -39.033 | 0.006** | 1,169.509** | 898.333* | 684.581* | -0.742** |
| | (85.342) | (0.002) | (579.2) | (541.044) | (371.77) | (0.355) |
| Audit Policy Controls | yes | yes | yes | yes | yes | yes |
| ***Panel D: Audits by IRA*** | Income | Share of evasion | Tot. taxable revenues | VAT taxable turnover | Operating costs | Operating costs/NPV |
| Audited at $t$ | -1.053 | 0.002 | 141.499 | -94.774 | 112.603 | 0.053 |
| | (31.884) | (0.001) | (262.49) | (145.493) | (109.159) | (0.129) |
| Audit Policy Controls | yes | yes | yes | yes | yes | yes |

*Notes.* Each cell displays estimates of separate OLS regressions in which the dependent variable is the variable named in the heading column by province and sector, and the independent variable is displayed in the row. Standard errors are clustered at the province and sector level (in parentheses). NPV stands for net present value. Audit policy controls include the characteristics of the tax filing, taxpayer, and TAs at the time of the tax filing, sector and province fixed effects, year of filing fixed effects, and age of tax filing fixed effects. *, **,*** denote statistical significance at the 10, 5, 1 percent level.

Table 4
Summary Statistics - TAs

*A. TAs: characteristics of the clients*

| N. TAs: 107,069 | mean | median | sd | 10th pct | 90th pct |
|---|---|---|---|---|---|
| N. taxpayers per TA | 31.13 | 17.57 | 106.51 | 2 | 64 |
| % clients in the same municipality | 61.65 | 61.29 | 23.14 | 30.50 | 96.28 |
| % clients in the same province | 89.83 | 93.77 | 11.97 | 73.13 | 100 |
| % clients audited | 4.81 | 3.21 | 7.28 | 0 | 11.54 |
| % evaders on clients | 3.11 | 1.35 | 5.65 | 0 | 8.11 |
| % evaders \| audit | 64.31 | 66.67 | 34.31 | 0 | 100 |
| Average evasion of clients | 0.32 | 0.30 | 0.25 | 0 | 0.65 |
| Sectoral specialization | 6.25 | 6.20 | 3.48 | 1.50 | 11 |
| % of new TAs in a year | 5.14 | 5.19 | 0.72 | 4.03 | 6.21 |
| % of closing TAs year | 3.67 | 3.57 | 0.34 | 3.35 | 4.22 |

*B. TAs: information on own tax filings*

N. TAs: 76,376; N. tax filings: 360,302
N. audited tax filings: 9,154; audited TAs: 6,410
N. TAs audited at least once: 8.39%

| N. TAs with positive evasion \| audit: 59.39% | mean | median | sd | 10th pct | 90th pct |
|---|---|---|---|---|---|
| Woman | 0.27 | 0 | 0.45 | 0 | 1 |
| Married | 0.72 | 1 | 0.45 | 0 | 1 |
| Age | 49.44 | 48 | 10.26 | 37 | 65 |
| Experience (years of activity) | 18.01 | 18 | 8.72 | 6 | 30 |
| N. employees | 0.92 | 0 | 1.83 | 0 | 3 |
| *Professional training* | | | | | |
| Dottori commercialisti | 0.45 | 0 | 0.50 | 0 | 1 |
| Ragionieri e periti commerciali | 0.31 | 0 | 0.46 | 0 | 1 |
| Revisori contabili | 0.10 | 0 | 0.30 | 0 | 1 |
| Consulenti del lavoro | 0.10 | 0 | 0.30 | 0 | 0 |
| Filed income | 42,558.20 | 28,065 | 57,279.47 | 5,500 | 89,581 |
| Profitability | 0.33 | 0.51 | 0.46 | 0.15 | 0.82 |
| Tax avoidance | 0.13 | 0.10 | 0.14 | 0 | 0.27 |
| Yearly % not congruent tax filings | 15.14 | 14.12 | 3.82 | 11.71 | 22.66 |
| Yearly % not coherent tax filings | 22.55 | 20.59 | 6.04 | 17.07 | 33.79 |
| Yearly % audited tax filings | 2.63 | 2.91 | 1.87 | 0.32 | 4.67 |
| Age of audited tax filings | 3.86 | 4 | 0.97 | 2 | 5 |
| Filed income \| audit | 54,936.71 | 34,768 | 71,862.19 | 5,175 | 118,916 |
| Evaded income \| audit | 18,760.68 | 1,996 | 107,954.95 | 0 | 41,447 |
| Filed income \| positive evasion | 55,351.64 | 34,909 | 75,050.452 | 4,968 | 119,259 |
| Evaded income \| positive evasion | 34,600.93 | 11,537 | 143,373.52 | 2,105 | 63,986 |
| Share of evasion on total income \| audit | 0.19 | 0.05 | 0.28 | 0 | 0.64 |
| Share of evasion on total income\| positive evasion | 0.34 | 0.25 | 0.29 | 0.04 | 0.83 |
| Appeal \| audit | 0.27 | 0 | 0.44 | 0 | 1 |

*Notes.* Income figures are expressed in euros.

Table 5

TA Characteristics by Type of Income Misreporting

| | Misreporting< 0 | | Misreporting> 0 | | $F$ test |
|---|---|---|---|---|---|
| | coef. | s.e. | coef. | s.e. | $p$-value |
| TA characteristics | | | | | |
| Own evasion*** | 0.163 | (0.012) | 0.237 | (0.003) | *0.000* |
| Own evaded income (euros)*** | 16,955 | (2,403) | 45,386 | (2,002) | *0.000* |
| Evader*** | 0.041 | (0.003) | 0.050 | (0.001) | *0.002* |
| N. audits*** | 0.243 | (0.009) | 0.334 | (0.003) | *0.000* |
| Evasion of clients*** | 0.294 | (0.004) | 0.356 | (0.001) | *0.000* |
| N. audits on clients*** | 3.438 | (0.064) | 4.460 | (0.020) | *0.000* |
| Experience | 19.110 | (0.113) | 19.027 | (0.026) | *0.470* |
| Profitability | 0.459 | (0.003) | 0.382 | (0.058) | *0.190* |
| Sectoral specialization | 9.460 | (0.039) | 9.409 | (0.009) | *0.209* |

*Notes.* This table reports OLS estimates of regressions of the characteristics of the TA on dummy indicators of the type of misreporting detected during an audit. The indicator of no misreporting is included but not reported, the constant is excluded. Robust standard errors are reported in parentheses. The sample includes all audited tax filings. The last column reports the $p$-value of the $F$ test of equality of coefficients for positive and negative misreporting. The statistical significance of the test is reported after the name of the corresponding dependent variables; *, **,*** denote statistical significance at the 10, 5, 1 percent level.

Table 6
Tax Evasion Spillovers

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| TA: Evasion other clients | 0.118*** | 0.116*** | 0.163*** | 0.143*** | |
| | (0.005) | (0.005) | (0.009) | (0.006) | |
| TA: Own evasion | | | | | 0.059*** |
| | | | | | (0.011) |
| Audit policy controls own | | | | | |
| Woman entrepreneur | 0.003 | 0.003 | 0.003 | 0.008*** | 0.007 |
| | (0.002) | (0.002) | (0.002) | (0.002) | (0.006) |
| Married entrepreneur | -0.026*** | -0.026*** | -0.027*** | -0.025*** | -0.033*** |
| | (0.001) | (0.001) | (0.002) | (0.001) | (0.005) |
| Entrepreneur 31-50 y.o. | -0.013*** | -0.013*** | -0.011*** | -0.014*** | -0.008 |
| | (0.003) | (0.003) | (0.004) | (0.003) | (0.009) |
| Entrepreneur >50 y.o. | -0.013*** | -0.013*** | -0.007* | -0.014*** | -0.004 |
| | (0.003) | (0.003) | (0.004) | (0.003) | (0.010) |
| Experience | -0.001*** | -0.001*** | -0.001*** | -0.002*** | -0.001*** |
| | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) |
| Firm size: 1-5 | -0.018*** | -0.018*** | -0.017*** | -0.015*** | -0.009* |
| | (0.002) | (0.002) | (0.002) | (0.002) | (0.005) |
| Firm size: 6-10 | -0.013*** | -0.013*** | -0.015*** | -0.007** | -0.017* |
| | (0.003) | (0.003) | (0.004) | (0.003) | (0.010) |
| Firm size: 11-15 | -0.017*** | -0.017*** | -0.018*** | -0.009* | -0.021 |
| | (0.005) | (0.005) | (0.007) | (0.005) | (0.016) |
| Firm size: 16-20 | -0.020** | -0.019** | -0.010 | -0.008 | -0.017 |
| | (0.008) | (0.008) | (0.010) | (0.008) | (0.024) |
| Firm size: >20 | -0.031*** | -0.030*** | -0.027*** | -0.018** | -0.029 |
| | (0.007) | (0.007) | (0.010) | (0.007) | (0.023) |
| Congruent | -0.099*** | -0.099*** | -0.097*** | -0.106*** | -0.095*** |
| | (0.002) | (0.002) | (0.002) | (0.002) | (0.005) |
| Coherent | -0.090*** | -0.090*** | -0.089*** | -0.100*** | -0.087*** |
| | (0.002) | (0.002) | (0.002) | (0.002) | (0.005) |
| TA: n. of clients/1000 | 0.002*** | 0.002*** | 0.001*** | 0.002*** | -0.088 |
| | (0.000) | (0.000) | (0.000) | (0.000) | (0.069) |
| TA: n. of provinces | 0.001** | 0.001** | 0.001* | 0.001** | -0.000 |
| | (0.000) | (0.000) | (0.000) | (0.000) | (0.001) |
| Year of filing FE | yes | yes | yes | yes | yes |
| Age of filing FE | yes | yes | yes | yes | yes |
| Sector FE | yes | yes | yes | yes | yes |
| Municipality FE | yes | yes | yes | | yes |
| Province FE | | | | yes | |
| IRA Office FE | yes | yes | yes | | yes |
| TA characteristics | | yes | yes | yes | yes |
| Audit policy controls peers | yes | yes | yes | yes | |
| Audit policy controls TA | | | | | yes |
| Sample: Large TAs | | | yes | | |
| R-squared | 0.186 | 0.186 | 0.194 | 0.067 | 0.278 |
| N. observations | 331,506 | 331,506 | 198,341 | 333,107 | 36,979 |

*Notes.* Columns 1 - 3 and 5 report OLS estimates of model (1), column 4 reports marginal effects of fractional probit model estimates. Standard errors are clustered at the TA level (in parentheses). The dependent variable is the own evasion of the taxpayer. The sample includes all audited tax filings of clients of TAs with at least another audited client. In column 3, the sample is reduced to tax filings of clients of TAs with at least 50 clients. Audit policy controls at the level of peers are computed as means of the variables listed in Table 3 for the audited clients. To avoid the incidental parameter problem, column 4 includes fixed effects for sectors at the 2-digit code. The reported measure of fit for the fractional probit is the pseudo R-squared. Baseline categories: age of entrepreneur: 30 years or younger, firm size: no employee. *, **,*** denote statistical significance at the 10, 5, 1 percent level.

Table 7
Tax Evasion Spillovers and TA Characteristics

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| TA: Own evasion × Experience T1 | 0.044** | | | | |
| | (0.021) | | | | |
| TA: Own evasion × Experience T2 | 0.064*** | | | | |
| | (0.015) | | | | |
| TA: Own evasion × Experience T3 | 0.066*** | | | | |
| | (0.018) | | | | |
| TA: Own evasion × Profitability T1 | | 0.062*** | | | |
| | | (0.012) | | | |
| TA: Own evasion × Profitability T2 | | 0.042** | | | |
| | | (0.019) | | | |
| TA: Own evasion × Profitability T3 | | 0.063*** | | | |
| | | (0.021) | | | |
| TA: Own evasion × Sectoral specialization T1 | | | 0.094*** | | |
| | | | (0.028) | | |
| TA: Own evasion × Sectoral specialization T2 | | | 0.046*** | | |
| | | | (0.018) | | |
| TA: Own evasion × Sectoral specialization T3 | | | 0.060*** | | |
| | | | (0.012) | | |
| TA: Own evasion × N. clients T1 | | | | 0.083** | |
| | | | | (0.034) | |
| TA: Own evasion × N. clients T2 | | | | 0.043** | |
| | | | | (0.017) | |
| TA: Own evasion × N. clients T3 | | | | 0.062*** | |
| | | | | (0.012) | |
| TA: Own evasion × Dottori commercialisti | | | | | 0.073*** |
| | | | | | (0.019) |
| TA: Own evasion × Ragionieri e periti commerciali | | | | | 0.060*** |
| | | | | | (0.018) |
| TA: Own evasion × Revisori contabili e periti | | | | | 0.062** |
| | | | | | (0.029) |
| TA: Own evasion × Consulenti del lavoro | | | | | 0.038 |
| | | | | | (0.026) |
| TA characteristics | yes | yes | yes | yes | yes |
| Audit policy controls own | yes | yes | yes | yes | yes |
| Audit policy controls TA | yes | yes | yes | yes | yes |
| *F* test of coefficients' equality: *p*-value | *0.642* | *0.538* | *0.251* | *0.400* | *0.742* |
| R-squared | 0.287 | 0.278 | 0.278 | 0.278 | 0.278 |
| N. observations | 32,492 | 36,979 | 36,979 | 36,979 | 36,979 |

*Notes.* This table reports OLS estimates with standard errors clustered at the TA level (in parentheses). The dependent variable is the own evasion of the taxpayer. The sample includes all audited tax filings of clients of TAs with at least another audited client and with observable TA audits. The variables denoted by T1, T2, and T3 are dummy variables with value 1 if the TA is respectively in the first, second and third tercile of the variable distribution. In column 3, the vector of TA characteristics includes sectoral specialization. Audit policy controls for own audit include the mean characteristics listed in Table 3 of the tax filings audited in the previous year. *, **,*** denote statistical significance at the 10, 5, 1 percent level.

## Table 8
## Information Channel

| | (1) | (2) | (3) |
|---|---|---|---|
| Peer audit at t-1 | 0.018*** | 0.013*** | 0.003 |
| | (0.003) | (0.003) | (0.003) |
| Own audit at t-1 | | 0.075*** | -0.003 |
| | | (0.004) | (0.006) |
| Time-varying characteristics | | | |
| Married entrepreneur | 0.353*** | 0.301*** | 0.230*** |
| | (0.007) | (0.007) | (0.009) |
| Age of entrepreneur | -0.114*** | -0.113*** | -0.077*** |
| | (0.002) | (0.002) | (0.002) |
| Years of activity | 0.101*** | 0.072*** | 0.078*** |
| | (0.002) | (0.001) | (0.002) |
| Size of the firm | 0.025*** | 0.016*** | 0.013*** |
| | (0.002) | (0.001) | (0.001) |
| TA: N. clients/1000 | -0.020*** | -0.021*** | -0.009*** |
| | (0.007) | (0.005) | (0.001) |
| TA: N. of provinces | 0.001 | 0.001 | 0.001** |
| | (0.001) | (0.001) | (0.000) |
| TA characteristics | yes | yes | yes |
| Year of filing FE | yes | yes | yes |
| Taxpayer FE | yes | yes | yes |
| Audit policy controls peer | yes | yes | yes |
| Audit policy controls own | | yes | yes |
| R-squared | 0.679 | 0.686 | 0.734 |
| N. observations | 15,921,793 | 13,928,480 | 7,666,069 |

*Notes.* This table reports OLS estimates of the model $Income_{i(j),t} = \gamma PeerAudit_{i(j),t-1} + \delta OwnAudit_{i,t-1} + \beta_1 z_{i,t} + \beta_2 z_{j,t} + \theta_i + \phi_t + \epsilon_{i(j),t}$, where $i$ and $j$ denote the taxpayer and the TA, respectively; $z$ is the vector capturing their time-varying characteristics of the taxpayer and the TA in the year of filing; and $\theta_i$ and $\phi_t$ are taxpayer and year fixed effects. Standard errors are clustered at the TA level (in parentheses). In columns 1 and 2, the dependent variable is the logarithm of the taxable income produced at t and reported at t+1, in column 3 it is the logarithm of the taxable income produced at t-3 and reported at t-2. Audit policy controls for peer and own audit include the mean characteristics listed in Table 3 of the tax filings audited in the previous year. *, **,*** denote statistical significance at the 10, 5, 1 percent level.

Table 9
Information Channel - Peers *vs* TA

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Peer audit same cluster | 0.013*** | 0.013*** | | |
| | (0.003) | (0.003) | | |
| Peer audit other cluster | 0.013*** | 0.013*** | | |
| | (0.003) | (0.003) | | |
| Non-peer audit same cluster | | 0.002 | | |
| | | (0.006) | | |
| Peer audit at t-1 | | | 0.013*** | 0.013*** |
| | | | (0.003) | (0.003) |
| Peer in the future audit at t-1 | | | -0.003 | |
| | | | (0.007) | |
| Peer in the past audit at t-1 | | | | -0.004 |
| | | | | (0.005) |
| Own audit at t-1 | 0.069*** | 0.069*** | 0.075*** | 0.075*** |
| | (0.004) | (0.004) | (0.004) | (0.004) |
| Taxpayer FE | yes | yes | yes | yes |
| Time-varying characteristics | yes | yes | yes | yes |
| Year of filing FE | yes | yes | yes | yes |
| Audit policy controls peer | yes | yes | yes | yes |
| Audit policy controls non-peer same cluster | | yes | | |
| Audit policy controls peer in the future | | | yes | |
| Audit policy controls peer in the past | | | | yes |
| Audit policy controls own | yes | yes | yes | yes |
| $F$ test of coefficients' equality: $p$ -value | *0.864* | *0.857* | | |
| R-squared | 0.686 | 0.686 | 0.686 | 0.686 |
| N. observations | 13,928,480 | 13,928,480 | 13,928,480 | 13,928,480 |

*Notes.* This table reports OLS estimates with standard errors clustered at the TA level (in parentheses). The dependent variable is the logarithm of the taxable income produced at t and reported at t+1. A cluster is defined by a $k$-means clustering algorithm over observable characteristics of the tax filing as described in Section 5.1. The number of clusters in each province is 100. "Peer audit same cluster" is a dummy variable with value 1 if in the previous year another client of the *same* TA and in the same cluster of the taxpayer received an audit. "Peer audit other cluster" is a dummy variable with value 1 if in the previous year another client of the same TA and in a different cluster of the taxpayer received an audit. "Non-peer" denotes clients of a *different* TA. Time-varying characteristics of the taxpayer and the TA in the year of filing are added. Audit policy controls for peer, non-peer, and own audit include the mean characteristics listed in Table 3 of the tax filings audited in the previous year. *, **,*** denote statistical significance at the 10, 5, 1 percent level.

Table 10
Memory of Information

| | (1) | (2) | (3) |
|---|---|---|---|
| Peer audit at t-1 | 0.015*** | 0.022*** | 0.006* |
| | (0.003) | (0.004) | (0.003) |
| Peer audit at t-2 | | 0.034*** | 0.006* |
| | | (0.004) | (0.004) |
| Peer audit at t-3 | | 0.030*** | 0.007** |
| | | (0.004) | (0.004) |
| Own audit at t-1 | 0.080*** | 0.079*** | 0.078*** |
| | (0.005) | (0.006) | (0.006) |
| Own audit at t-2 | 0.044*** | 0.046*** | 0.026*** |
| | (0.005) | (0.006) | (0.006) |
| Own audit at t-3 | 0.027*** | 0.026*** | 0.014** |
| | (0.005) | (0.005) | (0.005) |
| Taxpayer FE | yes | yes | yes |
| Time-varying characteristics | yes | yes | yes |
| Year of filing FE | yes | yes | yes |
| Audit policy controls peer at t-1, t-2, t-3 | | | yes |
| Audit policy controls own at t-1, t-2, t-3 | | | yes |
| $F$ test of $\beta_t^{own}$ equality: $p$-value | *0.000* | *0.000* | *0.000* |
| $F$ test of $\beta_t^{peer}$ equality: $p$-value | | *0.023* | *0.933* |
| $F$ test of $\beta_{t-2}^{peer} = \beta_{t-3}^{peer}$: $p$-value | | *0.390* | *0.738* |
| R-squared | 0.740 | 0.744 | 0.746 |
| N. observations | 7,526,420 | 7,044,423 | 7,044,423 |

*Notes.* This table reports OLS estimates with standard errors clustered at the TA level (in parentheses). The dependent variable is the logarithm of the taxable income produced at t and reported at t+1. The sample includes all tax filings of taxpayers filing in four consecutive years. Time-varying characteristics of the taxpayer and the TA in the year of filing are added. Audit policy controls for peer and own audit include the mean characteristics listed in Table 3 of the tax filings audited in the previous years. *, **,*** denote statistical significance at the 10, 5, 1 percent level.

Table 11
Sorting Channel

| Dep. Var.: Evasion of mover before move | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| New TA: Evasion of clients before move | 0.042*** | 0.048*** | 0.048*** | 0.062*** | 0.054*** |
| | (0.010) | (0.010) | (0.010) | (0.017) | (0.009) |
| New TA: Evasion of clients before move × Closure | | -0.024** | | | |
| | | (0.012) | | | |
| New TA: Evasion of clients before move × Takeover | | | -0.026 | | |
| | | | (0.018) | | |
| New TA: Evasion of clients before move × Old TA died or retired | | | -0.029 | | |
| | | | (0.026) | | |
| New TA: Evasion of clients before move × Other closure | | | -0.010 | | |
| | | | (0.015) | | |
| Old TA characteristics before move | yes | yes | yes | yes | yes |
| Year of move FE | yes | yes | yes | yes | yes |
| Audit policy controls mover | yes | yes | yes | yes | yes |
| Audit policy controls clients new TA | yes | yes | yes | yes | yes |
| Sample: Large TAs | | | | yes | |
| $F$ test of closure coefficients' equality: $p$-value | | | *0.702* | | |
| R-squared | 0.283 | 0.283 | 0.283 | 0.334 | 0.056 |
| N. observations | 30,330 | 30,330 | 30,330 | 16,441 | 32,187 |

*Notes.* Columns 1 - 4 report OLS estimates, column 5 reports marginal effects of fractional probit model estimates. Standard errors are clustered at the TA level (in parentheses). The dependent variable is the evasion of a mover before moving to a new TA. The sample includes taxpayers who changed TA at least once and were audited at least once before the move. Audit policy controls mover are the means of the variables listed in Table 3 of the audited tax filings compiled by the mover before the move. Audit policy controls new TA are computed as means of the same variables of the audited tax filings of clients compiled before the move. To avoid the incidental parameter problem, column 5 includes fixed effects for location at the province level, fixed effects for sectors at the 2-digit code, and excludes fixed effects for IRA office. The reported measure of fit for the fractional probit is the pseudo R-squared. *, **,*** denote statistical significance at the 10, 5, 1 percent level.

Table 12
Sorting and TA Specialization

| | (1) | (2) | (3) |
|---|---|---|---|
| New TA: Evasion of clients before move | 0.042*** | 0.042*** | 0.042*** |
| | (0.010) | (0.010) | (0.010) |
| New TA: Modal sector in mover's sector | 0.010 | | |
| | (0.007) | | |
| Old and new TAs same modal sector | | -0.007 | |
| | | (0.006) | |
| Old and new TAs same professional training | | | -0.006 |
| | | | (0.007) |
| Old TA characteristics before move | yes | yes | yes |
| Year of move FE | yes | yes | yes |
| Audit policy controls mover | yes | yes | yes |
| Audit policy controls clients new TA clients | yes | yes | yes |
| R-squared | 0.283 | 0.283 | 0.283 |
| N. observations | 30,330 | 30,330 | 30,330 |

*Notes.* This table reports OLS estimates with standard errors clustered at the TA level (in parentheses). The dependent variable is the evasion of a mover before moving to a new TA. The sample includes taxpayers who changed TA at least once and were audited at least once before the move. Audit policy controls mover are the means of the variables listed in Table 3 of the audited tax filings compiled by the mover before the move. Audit policy controls new TA are computed as means of the same variables of the audited tax filings of clients compiled before the move. *, **,*** denote statistical significance at the 10, 5, 1 percent level.

## Table 13
## Sorting along TA Own Evasion

| Dep. Var.: Evasion of mover before move | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| New TA: Own evasion before move | 0.079** | 0.078** | 0.079** | 0.078** |
| | (0.040) | (0.039) | (0.040) | (0.039) |
| New TA: Modal sector in mover's sector | | 0.022 | | |
| | | (0.034) | | |
| Old and new TAs same modal sector | | | 0.022 | |
| | | | (0.031) | |
| Old and new TAs same professional training | | | | -0.018 |
| | | | | (0.029) |
| Old TA characteristics before move | yes | yes | yes | yes |
| Year of move FE | yes | yes | yes | yes |
| Audit policy controls mover | yes | yes | yes | yes |
| Audit policy controls new TA own | yes | yes | yes | yes |
| R-squared | 0.435 | 0.435 | 0.435 | 0.435 |
| N. observations | 1,634 | 1,634 | 1,634 | 1,634 |

*Notes.* This table reports OLS estimates with standard errors clustered at the TA level (in parentheses). The dependent variable is the evasion of a mover before moving to a new TA. The sample includes taxpayers who changed TA at least once and were audited at least once before the move. Audit policy controls mover are the means of the variables listed in Table 3 of the audited tax filings compiled by the mover before the move. Audit policy controls new TA own are computed as means of the same variables of the audited tax filings compiled by the new TA before the taxpayer's move. To avoid the incidental parameter problem, fixed effects for location are included at the province level. *, **,*** denote statistical significance at the 10, 5, 1 percent level.

## Table 14
## Audit Policy Design

| | (1) | (2) | (3) |
|---|---|---|---|
| Peer audited before filing | 0.00055 | | 0.00292 |
| | (0.00034) | | (0.00418) |
| Peer detected evader before filing | 0.00010 | | -0.00047 |
| | (0.00026) | | (0.00077) |
| TA detected evader before filing | | -0.00009 | -0.00009 |
| | | (0.00025) | (0.00025) |
| Audit policy controls taxpayer | yes | yes | yes |
| Pseudo R-squared | 0.054 | 0.092 | 0.095 |
| N. observations | 8,780,364 | 97,148 | 97,148 |

*Notes.* This table reports marginal effects of probit models estimates with standard errors clustered at the TA level (in parentheses). The dependent variable is a dummy variable with value 1 if a tax filing is audited at t. The sample includes tax filings of income produced in 2012 and 2013. Audit policy controls taxpayer are computed as means of the variables listed in Table 3 for the audited tax filings.

## Table A1
## Information Channel - Spillovers Across TAs

| | (1) | (2) |
|---|---|---|
| Peer audit at t-1 | 0.013*** | 0.013*** |
| | (0.003) | (0.003) |
| Non-peer audit TA same cluster | 0.003 | |
| | (0.006) | |
| Non-peer audit same cluster TA same cluster | | 0.010 |
| | | (0.015) |
| Non-peer audit same cluster TA other cluster | | -0.007 |
| | | (0.014) |
| Own audit at t-1 | 0.075*** | 0.075*** |
| | (0.004) | (0.004) |
| Taxpayer FE | yes | yes |
| Time-varying characteristics | yes | yes |
| Year of filing FE | yes | yes |
| Audit policy controls peer | yes | yes |
| Audit policy controls non-peer | yes | yes |
| Audit policy controls own | yes | yes |
| R-squared | 0.686 | 0.686 |
| N. observations | 13,928,480 | 13,928,480 |

*Notes.* This table reports OLS estimates with standard errors clustered at the TA level (in parentheses). The dependent variable is the logarithm of the taxable income produced at t and reported at t+1. A TA cluster is defined by a *k*-means algorithm over observable characteristics of the TA's clients as described in Section 4. The number of different taxpayers' and TAs' clusters in each province is 100 and 20, respectively. "Non-peer audit same cluster TA same (other) cluster" is a dummy variable with value 1 if in the previous year a client of a *different* TA in the same cluster of the taxpayer and advised by a TA in the same (other) cluster of the taxpayer's TA received an audit. Time-varying characteristics of the taxpayer and the TA in the year of filing are added. Audit policy controls for peer, non-peer, and own audit include the mean characteristics listed in Table 3 of the tax filings audited in the previous year. *, **,*** denote statistical significance at the 10, 5, 1 percent level.