Dottorato di ricerca in Economia

Ciclo XXIX

S.S.D: SECS-P/03

Essays on tax compliance

Coordinatore: Ch.mo Prof. Gianluca Femminis

Supervisor: Alessandro Santoro

Enrico Urpis

Matricola: 4212550

Anno Accademico 2015/2016

Co-Autorship Disclaimer

Chapter 1 is a joint work with Prof. Alessandro Santoro, DEMS, Università degli studi di Milano-Bicocca.

Contents

Acknowledgement **vii**

Foreword **viii**

0.1 The low level of tax compliance in Italy ix

0.2 Estimates of tax compliance xi

0.3 First chapter xii

0.4 Second chapter xiii

1 The Impact of Audits on Tax Compliance: Evidence from Italy 1

1.1 Introduction 1

1.2 Some theoretical considerations 6

1.2.1 Dynamic and non-utility-based models 11

1.3 Institutional background 14

1.3.1 Definitions of base, audit, and after-audit years 16

1.3.2 Types of audits 19

1.4 Data description 19

1.4.1 Taxpayer characteristics 20

1.4.2 Panel characteristics 22

1.5 The challenge of non-random audit data 26

1.6 Identification strategy 29

1.7 Matching 32
1.7.1 Estimation of the propensity score 32
1.8 Discussion of the empirical approach 35
1.9 Results 38
1.9.1 Choosing matching algorithm and assessing matching quality 38
1.9.2 Estimates 41
1.9.3 Robustness using other matching approaches 42
1.10 Concluding remarks and directions for future research 45
Appendix 1.A The use of true income 46
1.A.1 The empirical approach with the use of the "true" income 48
Appendix 1.B Complete results for the other matching approaches . . 52

2 A Minimum Tax in Italy: Impact on Revenues and Redistribution 55

2.1 Introduction 55
2.2 Literature review 57
2.3 Theoretical background 61
2.3.1 Implementation of a minimum tax 61
2.3.2 The equity-efficiency trade-off 64
2.3.3 Taxpayers' reactions 65
2.4 Methods 69
2.5 Data 73
2.5.1 The database 73
2.5.2 Group 1 characteristics 77
2.6 Results 79
2.6.1 Outline 79
2.6.2 Differences in the income corrections 80

2.6.3 Differences in the correction 81
2.6.4 Effects of reactions to the minimum tax introduction . . . 86
2.6.5 A different representation 96
2.6.6 Trade-off between revenues and concentration 99
2.7 Conclusions 100
Appendix 2.A Groups classification 102
Appendix 2.B The two methodologies 107
2.B.1 General description 107
2.B.2 Presumptive statistical income estimation for Group G1B 108
2.B.3 Presumptive statistical income estimation for Group G1C 112
2.B.4 Italian Revenue Agency approach: Presumptive operative income 115
2.B.5 Istat methodology: Results for G1B 117
2.B.6 Istat methodology: Results for G1C 120
2.B.7 Istat methodology: Overall analysis of G1 and validation 124
2.B.8 RA presumptive operational income 128
2.B.9 Comparison of methodologies 130

Bibliography **135**

To Ching-Jung Chen

Acknowledgement

As always it is difficult to express the gratitude for all the persons who helped or were closed to me in doing such a big work. The biggest risk is to forget someone or to do not express the gratitude she deserves. For such reasons I would like to thank in chronological order who helped me directly and, for the one I forget to mention, please accept my apologies.

First of all, I would like to thank Professor Alessandro Santoro, who supported me in this process and constantly motivated me. He was really close guiding my work and his support was remarkable.

I have to express my sincere gratitude to my colleagues who studied with me in this process. Moreover, a special mention is deserved to Miss. Marina Bellani, Mr. Andrea Gurgone, Mrs Ayshe Alp and Miss. Ivana Ribac who helped me a lot in questioning my work and improving my exposition skills. I also need to thank doctor Raffaele Riccardi, doctor Laura J. Allevi, Mr Salvo Grasso, Mr Luca Zavarella and doctor Gianluca Di Fonzo their wise words helped me when I was in need. I also have to mention my director, Mr. Antonino Lucido, who supported me allowing a leave period that turned out to be really helpful.

As always, the parents and the family have to be acknowledge, thanks to their support I never gave up.

I should also express my gratitudes to my colleagues and managers that

helped me in various ways and support in difficult moments. The DEFAP staff for being present and their prompt replies. Families Masini, Grasso and Di Cioccio for hosting me when in need.

The final mention is for Miss Ching-Jung Chen, her support was decisive.

Foreword

0.1 The low level of tax compliance in Italy

Fulfilling fiscal duties is a compulsory obligation that all citizens of modern states are expected to pursue. By tax compliance we refer to the degree to which taxpayers comply with their fiscal duties. It is then crucial to investigate the extent to which taxpayers are loyal to the fiscal system. In fact, it is from the taxes that services are financed and they are so important that economists believe that civilization comes with taxation, as in Salanie (2002) .

However, not all fiscal duties are completely fulfilled by taxpayers. Every year a variable amount of taxes are not declared. Apart from possible mistakes, the reasons vary. They range from the act of not declaring marginal activities intended to be consumed just by the taxpayer or their family, to the intentional action of hiding the real taxable base. This intentional activity is performed in order to pay less than the due amount. Since paying taxes is costly, there is a consequent incentive to underreport them.

Taxes are essential for financing the services of states, so a decrease in resources can have harmful consequences for public life. For this reason, there has always been a need to understand why underreporting is happening, find a possible method to estimate it and correct this by having taxpayers be more compliant. A low level of compliance brings several issues. The first problem

concerns the decrease in the amount of resources that are collected. It is possible to see that a low level of compliance harms the government's ability to finance its activities, and so a low level of resources is associated with a low level of services and goods provided by the state. A high level of tax evasion also causes inequities and inefficiencies. One of the main tasks required of modern states is to redistribute resources among their citizens, but when the resources are hidden and not made available, this activity is compromised. Moreover, since not all taxpayers have the possibility of dodging their taxes, tax evasion affects the shares of the tax paid, meaning that some subjects pay a larger part of the fiscal burden than they should pay.

Italy is characterised by a low level of tax compliance and different attempts have been made to deal with this problem. These attempts have evolved over time, provided by new developments in information technology but also by a change in the approach to the problem. The instruments that have been displaced have been influenced by the evolution that has occurred in the literature. If in the past aim was stopping evasion from being profitable by setting an appropriate level of audits and fines (as in Allingham and Sandmo (1972)), more recent efforts have been devoted to increasing of tax compliance by promoting a more collaborative relationship (see the Italian parliamentary hearing May 2016) between the Italian Revenue Agency and taxpayers, since it is believed that taxpayers are not just driven by the desire to dodge their taxes if possible (as in Alm et al. (1992)).

Moreover, a change in the approach was required, since Italy continues to be characterized by a high level of tax evasion, as various estimations have pointed out.

0.2 Estimates of tax compliance

There are several different ways to estimate the level of tax compliance. The reason for the abundance of different methods in this field is the increasing number of available sources over the years (for example, the digitalization of information) and the possibility of analyzing this information using the latest innovations in informatics.

Different methodologies have been proposed throughout the years, and different estimations have been calculated. Estimations can be based on direct or indirect methodologies, can be pursued by public or private institutions, can follow a top-down or bottom-up approach, can be focused on the economy as a whole or just on particular sectors, etc.

Estimations of Italian tax evasion are provided by different institutes, but among them, the latest report of the Ministry of Economy and Finance provides a complete description of the problem of the low level of compliance in Italy.

Italy is still characterized by a high level of tax evasion. If the gap in all taxes as a whole is considered as an approximation of tax evasion (since this gap includes both intentional evasion and unintentional evasion, such as mistakes, misinterpretation of laws, missed payments, etc.), evasion is quite high, since this gap amounted to 23.6% in 2012 and 24.3% in 2013.

Moreover, there are differences among taxpayers. The same report points out that taxpayers with revenues lower than €30,987 are characterized by a very low level of tax compliance, since it is estimated that their rate of evasion is 51.56% higher than the average evasion rate.

Given such high levels of evasion, this work first investigates whether an audit has any effect on tax compliance at all. Secondly, it studies whether the implementation of a minimum tax can be a viable policy to solve the problem of the high level of tax evasion.

0.3 First chapter

The first chapter investigates the role of audits. Any tax system needs instruments to enforce fiscal laws. The most common instrument that has been used throughout the years is auditing the taxpayers. If a taxpayer is found not to be compliant, a fine is commuted.

This scheme can be considered to be the most successful because it is the most widespread. However, the effectiveness of audits has only recently been investigated. In more detail, according to DeBacker et al. (2015b), two opposite outcomes are possible, as taxpayers are influenced by two opposite effects: the target effect and the bomb-crater effect.

Under the target effect, audits can lower the level of evasion, since audited and fined taxpayers will change their expectations upward and declare more taxes to avoid incurring further audits. However, this effect may last for only a certain period of time, since, if no more audits take place, the expectation will shift downward. In the end, after an audit, a taxpayer's level of compliance will immediately increase, but it will later decrease to the previous level.

The bomb-crater effect is the total opposite. Some taxpayers will believe that they will be very unlikely to incur another audit immediately after a control. Since these taxpayers do not believe that an immediately subsequent audit is going to affect them, they will evade a great deal in the subsequent year(s) in order to recover from having to pay the fine.

This first chapter uses a database provided by the Italian Revenue Agency to shed some light on the behavior of Italian taxpayers. The focus is not just on which of the two effects is prevailing but is also on the magnitude of the effect and its extent in time. Using a combination of matching with difference-in-difference techniques, this work shows how in a particular context, such as Italy, audits can have a positive and lasting effect on tax compliance.

0.4 Second chapter

The second chapter explores the possibility of solving the problem of high Italian tax evasion by implementing a presumptive tax as a minimum tax. Presumptive taxes are used mostly, but not exclusively, in developing countries where information on the tax base is too difficult to obtain or too unreliable.

Presumptive taxes have been studied extensively from a theoretical point of view, and several authors have also provided descriptions of their implementations over time and in different countries. Presumptive taxes are investigated in terms of their pros and cons and issues of fairness and efficiency.

The aim of the second chapter is to shed some light on the trade-offs associated with the implementation of a minimum tax. An increase in the total amount of tax collected is associated with a decrease in differentiation across the shares of paid taxes. In more detail, this study first investigates the potential outcome of the implementation of a minimum tax without considering reactions from taxpayers.

However, the implementation of a new policy might cause a change in taxpayer behavior. Two different effects might take place: the leaving effect and the leveling effect. For some taxpayers, presumptive taxes might be too high, so they will decide to leave the market. Other taxpayers whose presumptive taxes are lower than their declared taxes might consider it more convenient to pay the presumptive tax and bear the risk of incurring an audit. Either of these effects might change the outcome and significantly affect the implementation of a minimum tax.

To investigate the implementation of the minimum tax and the effects of the reactions, two different specifications of presumptive taxes are used. This choice was made to provide more general results. Moreover, this work is focused on a very specific group of taxpayers that various studies consider to be characterized

by a low level of tax compliance.

The aim of this work is to provide estimates of both the trade-off between equity and efficiency associated with the implementation of a minimum tax and, later, the effects of different levels of reactions on the outcome. Moreover, this analysis is specifically carried out based on the same database adopted in the first chapter and, thus, provides specific results for the Italian context. In the end, this work shows the conditions for which the implementation of a minimum tax would be convenient in a context like that of Italy.

Chapter 1

The Impact of Audits on Tax Compliance: Evidence from Italy

Santoro, A.; Urpis, E.

1.1 Introduction

In all developed countries, revenue agencies devote significant efforts and public resources to auditing taxpayers' reports and sanctioning taxpayers if they are found guilty. Despite the magnitude of the administrative and compliance costs created by these activities, the literature on the actual impact of audits has developed only recently.

Andreoni et al. (1998) identify a conflict between experimental and econometric results, which they discuss in Section 6.4 of their paper. Experimental studies usually report that compliance increases significantly in later rounds of

experiments among those participants who were audited in previous rounds. On the contrary, audit studies, mainly based on Internal Revenue Service (IRS) data, find no significant effect of prior audits on reporting behavior. The authors offer two possible explanations for this lack of deterrence. First, audits in the real world are not always fully efficient, that is, they may fail to uncover existing noncompliance, so that taxpayers may conclude that it pays to cheat. Second, alternatively, when taxpayers feel that an audit was a bad experience, they may want to evade more in the future in an attempt to get back at the revenue agency. This second explanation paves the way for a negative, and somewhat counterintuitive, effect of audits on compliance. Significantly, Andreoni et al. (1998) conclude this discussion by writing that "more research is needed both to confirm whether there is any specific deterrent effect of an audit and to uncover the reasons for the presence or the absence of such an effect."

This gap has been filled by more recent literature only to a limited extent. According to DeBacker et al. (2015b), two opposing expectations about the aftereffects of tax audits have emerged.

The first, and perhaps more intuitive, expectation is that experiencing audits leads taxpayers to revise their perceived audit probabilities upward and therefore reduce their subsequent noncompliance. Afterward, for each year that they do not experience audits, taxpayers revise their perceived audit probabilities downward and, thus, increase their noncompliance. In that case, the post-audit tax payment trend would consist of an immediate increase followed by a decrease, which is called the *target effect* by (Hashimzade et al., 2014). This effect is conjectured and verified, using data from a natural experiment, by Kleven et al. (2011), who argue that, in the real world, experiencing an audit leads to an increase in the perceived probability of being *detected* since taxpayers are audited only if tax inspectors believe potential income can be uncovered.

More recent contributions that find results consistent with a prevailing target effect are Kleven et al. (2011), Advani et al. (2015), and DeBacker et al. (2015a). All of these three studies use random audits.

Intriguingly, the second expectation is the complete opposite. Taxpayers may, correctly or incorrectly, perceive that auditors rarely come back immediately after an audit and that it is safest to evade taxes right after an audit. As years pass, the risk that auditors will come back increases, and, thus, it is best to reduce noncompliance. In this case, the post-audit tax payment trend would be an immediate decrease followed by an increase. This short-term negative effect is related to the *bomb-crater effect* first identified by Guala and Mittone (2005) and repeatedly found in more recent lab experiments. DeBacker et al. (2015b), using data from real-world operational audits, find a similar pattern for corporations using IRS and financial statement data. Corporations gradually increase their tax aggressiveness for a few years following an audit and then reduce it sharply.

In this study, we contribute to this literature by focusing on the impact of audits on tax compliance by individuals in Italy. Our database consists of a panel of 528,540 VAT-registered taxpayers who earn income from self-employment and/or from sole proprietorships. This panel represents the universe of VAT-registered taxpayers who reside, for tax purposes, in three of the main regions of Italy, Lombardy, Lazio, and Sicily, which are located in the north, center, and south of Italy, respectively. The periods we observe include tax years 2007 and 2011, for a total of 2,642,700 observations. The panel is perfectly balanced, that is, it includes exclusively taxpayers who issue their tax returns for all of the five years. For the same tax years, we observe tax returns that are audited by the Italian Revenue Agency. More precisely, we observe a total of 31,135 audited tax declarations referring to tax years 2007 and 2008. Of these, 22,175 are single

audits conducted on taxpayers audited only once over the periods observed.

We believe our research can be of general interest for two reasons.

First, we use results from the real world, that is, operational audits, whereas the existing empirical literature uses random audits[1]. Although random audits clearly do offer an empirical advantage by providing a clear counterfactual scenario, we share the view of Slemrod (2016) that "the external validity of these results is somewhat problematic. Because taxpayers audited under NRP are informed that they have been randomly selected for research purposes, these audits may not have the same impact on the perceived probability of a future audit as an operational audit." This argument applies to DeBacker et al. (2015a) and to Advani et al. (2015), but not to Kleven et al. (2011). Another critical argument put forward by Slemrod (2016) is that, since revenue agencies do not select taxpayers randomly, taxpayers subject to random audits are not representative of those who are typically subject to audits, and their behavior may not be representative of that of taxpayers who are normally targeted for operational audits. These issues imply serious limitations to drawing conclusions about policy from the results based on random audits.

The second reason that this study is of general interest is that Italy differs from countries for which the impacts of tax audits have already been tested (i.e., Denmark, the UK, and the US) because of its low tax morale and relatively inefficient tax institutions. Both of these features may undermine the impact of tax audits by lowering the perceived probability of being sanctioned and/or the extended value of sanctions, which includes the social stigma arising from the application of sanctions.

Even if a taxpayer knows she is being targeted by tax authorities, she still may not increase her tax compliance if she believes that, in the future, she will

[1]The same database is used by Mazzolini et al. (2017), which is a companion study to this one

be able to find alternative ways to evade. Thus, an audit acts as a deterrent of future evasion to the extent that the threat of an audit in the future is credible, which, in turn, depends on the quality of enforcement institutions. Furthermore, tax morale and social norms can play an important role. The main determinants of tax morale are intrinsic motivations and perceptions of the fairness of the tax system and the legitimacy of the state Luttmer and Singhal (2014). Thus, even if a taxpayer does feel she is being targeted by authorities, she still may not increase her tax compliance when no social stigma is attached to evasion and/or if she feels justified in continuing to evade taxes raised by a state that she perceives as unfair or illegitimate.

Clearly, operational non-random audits pose an identification threat, since taxpayers selected for audit are likely to be different from those who are not audited. Thus, a simple comparison between tax returns issued by audited and non-audited taxpayers would be biased. We deal with this problem by using an identification strategy based on the use of a difference-in-differences matching estimator (see page 56 of Caliendo and Kopeinig (2008) and the literature cited therein), which combines propensity score matching with the difference-in-differences technique. Furthermore, we exploit the richness of our database by running some placebo tests.

Overall, our results indicate that compliance increases after audits, since income reported by audited taxpayers increases with respect to that reported by matched non-audited taxpayers. Moreover, in absolute terms, this increase seems to respond mainly to the lag between the time when the audit is conducted and the after-treatment tax year that is considered. More precisely, an audit conducted during tax year t has a smaller but positive effect on tax returns issued that year and referring to the previous year and a larger effect on tax returns issued during year $t+1$ and referring to tax year t. These results are

obtained for different audit years and post-treatment years. However, the only year in which an impact is expected but not found is tax year 2009, the year when the Italian economy was most strongly hit by the economic recession. This finding may indicate that, as expected, the impact of an audit depends also on the economic context, and, particularly, on liquidity constraints.

The paper is organized as follows. In Section 1.2, we suggest some theoretical reasons to justify the possible roles played by tax morale and the (in)efficiency of institutions. In Section 1.3, we describe some institutional features of the Italian tax system. In Section 1.4, we provide a description of the database along with some descriptive statistics. In Sections 1.5, 1.6 and 1.7, we describe our identification strategy, which we discuss critically in Section 1.8. Section 1.9 provides the result and is divided into two subsections. The first one includes a test assessing the matching quality, and the second one estimates the treatment effect and placebo regressions. Section 1.10 concludes.

1.2 Some theoretical considerations

In the context of the classical Allingham-Sandmo model (Allingham and Sandmo (1972), A-S hereinafter) of tax evasion, the occurrence of an audit can enter the process of updating beliefs.

Here, we assume that the model that applies to individuals can also be applied to small businesses, since, in our dataset, these small businesses include a significant amount of personal work and very limited amounts of external work and capital. We shall show these metrics in 1.4.1.

Moreover, the data reasons for adapting the A-S model to large businesses (see Shackelford et al. (2007)) do not apply in this case, as asymmetric information due to the presence of an agent is absent. Usually, managers of large businesses face a dilemma due to the presence of two different audiences, stake-

holders and the IRS. By reporting truthfully, managers may increase the value of their stocks, but, at the same time, they have to pay the required taxes. Alternatively, managers may not report total revenues in order to pay fewer taxes, but, in that case, they cause their stocks to depreciate.

In the context of small businesses, no such agency problems arise, so this dilemma is absent. Thus, dealing with small businesses does not require substantial modification to the A-S model, contrary to the case of evasion by large businesses.

Another difference is provided by Slemrod (2004), who specify that entrepreneurs and small businesses are characterized by a different risk approach than large enterprises are. Small firms and entrepreneurs are risk averse, whereas large companies are characterized by risk neutrality. The main reason is that small businesses are focused on just one activity, implying a difference between the cost of a dollar if caught and the benefit of a dollar gained by evading.

Moreover, large enterprises can invest significant resources in legally eluding the tax norms. Such investments are prohibitive for small firms, which must simply evade taxes if they want to reduce their tax rates.

According to the A-S model, if taxpayer i decides to evade taxes, her expected income is equal to

$$p[(1-\tau)Y^i - \theta\tau E^i] + (1-p)[(1-\tau)Y^i + \tau E^i] \tag{1.1}$$

where p is the perceived probability of an audit, $\tau > 0$ is the proportional tax rate, and $\theta > 0$ is the sanction rate. The taxpayer chooses E_i, the amount of true income that is not reported (thus, $E^i = Y^i - \widehat{Y}^i$, where $\widehat{Y}^i$ is reported income). The taxpayer decides to evade only if her expected income under evasion is higher than that under no evasion, in which case she receives $(1-\tau)Y^i$ with certainty. Then, it is immediate to see that if $p < 1/(1+\theta)$, the taxpayer will

evade her income completely, whereas, on the contrary, she will not evade if $p > 1/(1+\theta)$.

To allow for internal solutions (i.e., for optimal evasion between 0% and 100%) either risk aversion or a (convex) function of concealment costs can be inserted into the model, but the roles of the two fundamental parameters of deterrence, p and θ, are not altered.

Some simple extensions of the model clarify the roles of tax morale and of the quality of institutions in this process of updating beliefs. We are interested in analyzing these extensions since they provide support for the external validity of our research on Italian taxpayers.

Looking at expression (1.1) it is clear that p should be interpreted as the perceived probability that any evasion will be detected, which is not necessarily equal to the probability of being audited. As stated by (Kleven et al., 2011), it "is the product of the probability of being audited again and the probability of undeclared income being uncovered conditional on audit." Thus, we can write

$$p = p^u p^a \tag{1.2}$$

where p^u is the probability of undeclared income being uncovered in the case of an audit and p^a is the probability of being audited again.

Now, in the original A-S model, the revenue agency is assumed to be 100% efficient, so that $p^u = 1$ and p can be interpreted as the probability of an audit. However, this assumption is not realistic, since the revenue agency may be unable to detect all tax evasion.

As stressed by (Kleven et al., 2011), if a taxpayer is audited, she may revise her beliefs about p^a upward. However, the final impact on the taxpayer's decision depends on p^u, which, in turn, depends on the perceived efficiency of the

revenue agency[2].

The available literature on the impact of tax audits on subsequent compliance refers to countries where the quality of tax enforcement institutions is generally deemed to be high (i.e., Denmark, the US, and the UK). A recent survey by Oecd (2015) lists eight criteria to provide a high quality strategic approach for managing taxpayers' compliance. The tax administrations of Denmark and the UK satisfy all of them, and that of the US fails with respect to only one of these criteria (see Oecd (2015), Table 3.9, p. 129). These outcomes are consistent with the overall evaluation of government quality. Denmark ranks first and the UK ninth according to the European Quality of Government Index (EQI).

On the contrary, the Italian Tax Administration fails to satisfy at least five of the eight criteria mentioned above (see Oecd (2015), Table 3.9, p. 129) and, in particular, the debt collection function is inefficiently administered (Oecd (2016)). The inefficiency of the Italian tax administration relative to other OECD countries is just one aspect of the problem of low government quality, as Italy ranks only 24th among the 31 countries listed in the EQI ranking, although northern regions perform better than central and southern ones.

In sum, on an a priori basis, one could expect Italian taxpayers to perceive a low probability that evasion is detected, which means that p^u is low, undermining the possible increase of p^a generated by an audit.[3]

A very similar reasoning applies to the analysis of sanctions. The expected utility framework of the A-S model has been enriched by also including the role of social sanctions, such as the stigma attached to detected evasion, in the

[2]NEW PART point 2: This is the target effect that we shall illustrate in 1.2.1.

[3]Clearly, if the outcome of the audit, that is, the actual amount of evasion that is legally assessed, were fully observable, we could distinguish between cases where p^u is also updated and cases where it is not. For example, Mazzolini et al. (2017) distinguish between audits with a null outcome and audits where the taxpayer accepts, implicitly or explicitly, the assessment by the revenue agency, but they cannot observe the final outcomes of audits followed by a legal suit.

model. According to this view, the parameter θ depends on both the level of legal sanctions, θ^L, and that of social sanctions, θ^S, as follows

$$\theta = f(\theta^L, \theta^S) \tag{1.3}$$

with f strictly increasing in both arguments.

On the one hand, the *effective* value of θ^L varies across different countries because of heterogeneities within tax codes and also because of the frequency and magnitude of tax amnesties. In Italy, there has been a tax amnesty, in some form, every two years, on average, for the past twenty years. Thus, the expected value of θ^L is very low in Italy as compared to that in other countries.

On the other hand, countries for which evidence of the impact of audits is available can be assumed to have quite high values of θ^S.

The US ranks first and Denmark fourth in the ranking of tax morale provided by Alm and Torgler (2006) using the World Value Survey (see Alm and Torgler (2006), Figure 2, p. 239). On the contrary, although Italy has a medium tax morale rank (7th) on average according to Alm and Torgler (2006), it is generally believed that in some parts of the country, especially in the central and southern regions, tax morale is very low (Cannari and D'Alessio (2007)). This heterogeneity traces back to Italian history. As stressed by Alm and Torgler (2006), low tax morale can be explained by the historical conflicting relationship between secular and religious authorities that characterizes the center and south of Italy. Low tax morale is generally associated with limited social sanctions and stigma associated with tax evasion. The tax evader, in Italy, is seen as a smart person rather than as a law breaker or a criminal.

Thus, in Italy, both the values of θ^L and of θ^S are lower than those in many other developed countries, and the condition ensuring no evasion (i.e., $p > 1/(1+\theta)$) is less likely to be satisfied for any given p.

To sum up, there are reasons to expect that the impact of audits on a country such as Italy is different from that on countries such as the US, Denmark, and the UK.

1.2.1 Dynamic and non-utility-based models

The A-S model also presents an analysis of taxpayers' behavior in a dynamic context. If different periods of time are included in the analysis, the model conjectures that taxpayers will exhibit two opposite behaviors. Myopic individuals will not consider the effects of their actions in a multi-period context and will therefore simply behave as if every period of time is independent from the others. On the contrary, consistent subjects maximize their lifetime utilities and are aware of the fact that cheating in one period affects their positions in subsequent periods. More precisely, if each taxpayer found cheating is also investigated in all years prior to his last full payment, only the consistent individuals will use this information. In this setting, the authors illustrate that some taxpayers may be honest and some may dodge taxes given individual characteristics, as myopic individuals declare fewer taxes than consistent individuals.

The dynamic setting of the A-S model incorporates discrete time and an infinite life expectancy. In fact, consistent individuals are aware of the effect that cheating has in the current year, and, for this reason, their present decisions are influenced by their past decisions. This cohort of taxpayers is aware that an audit in the present day will bring controls for previously underreported income, so income not reported today will also influence the decision to report truthfully tomorrow.

The A-S dynamic context thus provides insight into the possible reasons for some taxpayers declaring more than would be expected in a single period setting. As consistent taxpayers are aware of the impact of their decisions in

time, they tend to comply more than myopic ones do. Thus, in this context, it is shortsighted to evade taxes.

However, Alm et al. (1992) find that taxpayers may not behave as profit maximizers. Alm et al. (1992) focus on tax morale by reproducing in a lab experiment the choices that individuals face in their real lives. The authors find that experimental subjects are driven by a multitude of forces, such as the over- or underestimation of the probability of being detected. Moreover, they also find that when the probability of being discovered cheating approaches zero, some subjects are still willing to participate in the tax system and declare thoughtfully.

Since the choice faced by the taxpayer is not strictly individualistic, Alm (2012) consider the network effects of the influence of other people on the behavior of taxpayers to further investigate this assumption. In the end, they show that if taxpayers suspect that the individuals that they observe cheat, these taxpayers will adapt and report fewer taxes. This phenomenon is called the peer effect to emphasize the role that one close person has on another.

The impact of the occurrence of an audit has become the central object of interest in behavioral models. It has been reported that the behavior of an individual can change when that individual is audited. Among other studies, Hashimzade et al. (2014) demonstrate that the effect of an audit on the following year's behavior is difficult to predict since two mechanisms are at work: the target effect and the bomb-crater effect. In their model, the target effect is personal to each taxpayer and is positively related to the control. If a person is audited, the target effect increases his compliance in the following years, since he has the experience of being targeted by the tax authority. At the same time, the bomb-crater effect means that being audited decreases the level of compliance, since the taxpayer believes that the chance of being audited two years in a row

is negligible. For this reason, the authors show that, at least in the short term, it is difficult to predict the effect of an audit, since these two effects work at the same time but in opposite directions.

More specifically, the bomb-crater effect is the focus of Guala and Mittone (2005) and Kastlunger et al. (2009). These authors explain this particular behavior by considering two possible reasons. First, some subjects tend to evade more immediately after being audited because of the loss repair effect, since an audited taxpayer might try to restore his losses by engaging in tax evasion in subsequent filings, as in Maciejovsky et al. (2007). The second reason is a misspecification of chances, as described in Maciejovsky et al. (2007) and in Kastlunger et al. (2009). The behavior of taxpayers may be similar to that of soldiers in the First World War. Taxpayers believe that it is improbable that they will to be audited again after being audited. Moreover, Mittone (2006) investigates the long-term effect of audits and finds that in experiments, the bomb-crater effect weakens after 30 rounds.

Since the analysis of dynamic models shows that either the bomb-crater effect or the target effect arises depending on certain conditions, it is interesting to analyze the response to an audit in a context like Italy, where the level of tax compliance is low. Previous studies have analyzed the responses to audits of taxpayers characterized by a high level of compliance, so comparing the responses of taxpayers with a lower level of compliance can show similarities or differences in behavior. This comparison can help to provide suggestions for better policies, which can be tailored to the environment and the characteristics of the taxpayers.

1.3 Institutional background

In Italy, individual taxpayers earning income from self-employment and/or from sole proprietorships are required to register for value-added tax (VAT) purposes. Both of these incomes are basically self-reported and are not subject to third-party information[4]. Tax returns are based on accounting books, but for many small economic activities, a simplified accounting regime is adopted[5], so that only some of the transactions relevant for tax purposes are recorded. Thus, a tax return can diverge from the accounting books to some extent if a simplified accounting regime is adopted. We shall refer to this possibility as *ex-post tax adjustment*.

The lack of third-party information implies that tax audits are a particularly important tool for reducing the evasion of taxes on these incomes.

In Italy, audits are the most important kind of tax verification activity conducted by Agenzia delle Entrate (AE), although some specific verification activities are carried out by Guardia di Finanza (a separate tax police body). Of AE's entire workforce of approximately 30,000 units, 12,000 are involved in audit activities. The whole auditing process consists of five different steps: i) mapping and planning; ii) risk assessment; iii) selection of taxpayers to be audited; iv) audits and assessment; and v) settlement or trial (in the case of a positive assessment). Steps i)-iv) are administered by central, regional, and provincial directorates.

More specifically, the central directorate (Direzione centrale dell'accertamento) is entirely in charge of step i), which is accomplished by setting specific targets in terms of the number and type of taxpayers (large, medium, and small companies and self-employed individuals) to be audited each year. Then, whereas

[4]Some types of income obtained by professionals working with firms are subject to a withholding tax.

[5]Under given thresholds, the simplified accounting regime is the default option, but the ordinary accounting regime option can be chosen by the taxpayer if she is willing to do so.

audits on large taxpayers are administered mostly at the central level, audits on medium (approximately 60,000 taxpayers) and small businesses and self-employed individuals (approximately 4 million taxpayers) are administered by regional and provincial directorates, respectively, in collaboration with the central directorate. This collaboration is particularly important within step ii) (risk assessment).

An important source of information for AE is provided by Business Sector Studies (Studi di Settore (SDS), see Santoro and Fiorio (2011)). For each taxpayer subject to an SDS, a presumptive turnover is computed by multiplying input values, as reported by the taxpayer, by input productivities, as computed by AE.

Taxpayers are classified as congruous or non-congruous if their reported turnovers are higher or lower, respectively, than the presumptive turnover. Taxpayers know their presumptive turnovers, and they know that they are more likely to be audited if they are non-congruous. However, congruous taxpayers can also be audited, and, conversely, the vast majority of non-congruous taxpayers are not audited. The whole range of these activities generates different lists of potential targeted taxpayers that are then transmitted to local directorates.

However, both regional and provincial directorates possess a remarkable degree of autonomy in the selection of the taxpayers to be audited (step iii). On the one hand, they can conduct their own risk assessments, based on information collected on the spot, and identify risky taxpayers not selected at the central level; these assessments are particularly important for the Italian economy, which is characterized by a number of locally specialized districts. On the other hand, since local directorates are in charge of the achievements of their targets, they can at least partly ignore the taxpayers included in the list provided by the central directorate. Although information on the identities of the

taxpayers who are actually audited is not publicly available, the nature of this process is such that it is reasonable to presume that the main selection criteria are the geographical location, sector, size, and legal type of the economic activity (see also Section 1.7.1)

1.3.1 Definitions of base, audit, and after-audit years

In Italy, individual taxpayers are required to pay taxes on all personal income earned in a given tax year. The tax year usually begins on the 1st of January and ends on the 31st of December. Incomes earned in a given tax year have to be reported between May and September of the following calendar year. Thus, incomes earned between January 1st and December 31st of year t have to be reported between May and September of year $t + 1$.

Once a tax report has been issued, it can be audited. However, audits are subject to an *expiration deadline*. In Italy, a tax report referring to year t must be audited by the 31st of December of year $t + 5$. After this time, evasion can be prosecuted only if it is associated with some other crime.

In this study, we should define the *base year* as the tax year in which the return is filed. In contrast, the *audit year* is the year in which the audit occurs. However, these definitions do not overlap exactly with the calendar year. According to the AE definition, a "year t" audit is an audit carried out between July 1st of year $t - 1$ and June 30th of year t.

Because of the expiration deadline mentioned above, tax returns that refer to base year 2007 can be audited in audit years between 2009 and 2013, whereas tax returns that refer to base year 2008 can be audited in audit years between 2010 and 2014.

Finally, we define as an *after-audit year* a tax year whose tax return can be influenced by an audit. In general, if an audit is conducted during year t, its

impact, if any, should be fully captured in tax returns that refer to year t, but it is possible that an audit can have some effects on tax returns that refer to previous years, namely, year $t-2$ and, in particular, year $t-1$.

To illustrate this possibility, Figure 1 provides a useful example where an audit is conducted in audit year t and tax returns that refer to tax years $t-2$ and $t-1$ are affected.

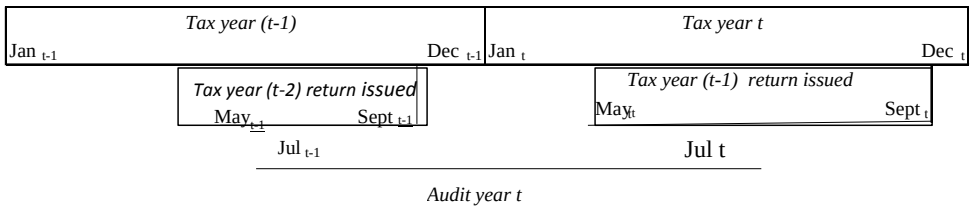


Figure 1.1: Time framework of tax returns and of audits

The first tax return we consider is that for year $t-2$. This return may be influenced by the audit if the audit is conducted between July and September of year t and only if the tax return was issued between May and June of the

same year. Even in that case, however, year $t-2$ is surely over when the audit is conducted, so the impact of the audit can only be reflected in ex-post tax adjustments. Overall, the possibility that an audit conducted in audit year t influences the tax return for base year $t-2$ is very unlikely, but it cannot be ignored on a priori grounds. No effect should be observed for the tax returns for years before $t-2$.

A tax return for year $t-1$ is more likely to be influenced by a year t audit. Unless the audit is conducted in June or July of year t and the tax return is issued in May of the same year, the tax return is issued after the audit is conducted. Moreover, if the audit is conducted between July and December of year $t-1$, accounting books have still to be completed. However, if the audit is conducted between January and July of year t, accounting books are completed, so the impact of an audit is reflected only by ex-post tax return adjustments.

Since the audit year extends from the 1st of July of year $t-1$ to the 30th of June of year t, if a return is audited on July 1, 2009, it means that the control is taking place in audit year 2010. Moreover, the taxpayer still has until September 30, 2009 to file a return for the year 2008. For this reason, it is possible that an audit in audit year 2010 can influence the return for 2008.

Because of these differences among calendar, report, audit, and post-audit years, the database includes the effects of audits occurring from 2011 to 2013.

Finally, from the tax return referring to tax year t onwards, the effects of the audit, if any, should be fully observable. Thus, to sum up, for every audit (treatment) conducted in year t, we can distinguish three sets of years:

- years from t onwards, which are *after-audit years* for which the effect of the audit (if any) should be fully observable;
- year $t-1$, which is an *after-audit year* for which the effect of the audit (if any) should be partially observable (i.e., only observable for ex-post tax

return adjustments);

- years $t - n$ for $n \geqslant 2$, which are *pre-audit years* for which the effect of the audit is dubious.

1.3.2 Types of audits

We define as *single* an audit conducted in a given year that inspects only one tax return issued by a given taxpayer. For a single audit, there is one audit year and one base year. We define as *multiple* an audit conducted in a given year that inspects more than one tax return issued by a given taxpayer. In particular, an audit conducted in January of year t could inspect tax returns referring to the base years from $t - 5$ to $t - 2$, inclusive.

On the other hand, using a panel, we can observe a taxpayer audited only once (with a single or multiple audit) or a taxpayer audited *repeatedly* (again, with single or multiple audits).

The distinction between these different types of audits is relevant for the purpose of identifying an audit's effects. Although, as we shall explain below, we are quite confident that we can identify some of the determinants of the probability of conducting a standard, single audit, the information used by the Revenue Agency to decide when to opt for a multiple audit or to audit a taxpayer more than once is not available to us.

1.4 Data description

Our database consists of a panel of 528,540 VAT-registered taxpayers. A VAT-registered individual taxpayer is a taxpayer who runs an economic activity on his own, as self-employed or as a single entrepreneur. Thus, his sales and purchases are, in principle, subject to the VAT.

1.4.1 Taxpayer characteristics

The taxpayers included in the analysis perform different kinds of activities. Most of them are active in services or commerce. As Table 1 shows, less than one tenth of these taxpayers are involved in agriculture and industry. Considering both 2007 and 2008, we observe only a few migrations across different sectors, so the choice of activity can be considered stable.

Table 1.1: Economic sectors of the taxpayers included in the database

	2007		2008	
Economic sector	n.	%	n.	%
Agriculture	37,370	7.07%	37,313	7.06%
Services	256,366	48.50%	256,477	48.53%
Commerce	134,370	25.42%	134,357	25.42%
Construction	66,892	12.66%	66,876	12.65%
Industry	33,542	6.35%	33,517	6.34%
Tot	528,540	100.00%	528,540	100.00%

We now present the features that justify the use of the A-S model. The taxpayers in the sample are running businesses for which the most important input is the workforce of the owner. The use of other's people work is marginal, and the same is true for the use of capital goods.

Three quarters of the individuals included in the database do not hire any workers at all. The remaining quarter typically hires fewer than ten workers in the observed years. The limited fraction of taxpayers hiring ten or more individuals hires an average of approximately fifteen workers in both years. Table 2 illustrates that this pool of taxpayers uses an extremely limited amount of the external workforce and that the entrepreneurs are actually involved in production.

Another characteristic of the sample is the limited amount of capital equip-

Table 1.2: Number of hired workers per activity

	2007		2008	
N. workers	n.	%	n.	%
0	398,029	75.31%	390,655	73.91 %
1	55,499	10.50%	57,955	10.97 %
2	27,748	5.25%	29,403	5.56 %
3	15,594	2.95%	16,711	3.16 %
4	9,386	1.78%	10,128	1.92 %
5	6,080	1.15%	6,482	1.23 %
6	4,097	0.78%	4,372	0.83 %
7	2,860	0.54%	3,016	0.57 %
8	2,024	0.38%	2,150	0.41 %
9	1,544	0.29%	1,653	0.31 %
10 +	5,679	1.07%	6,015	1.14 %
Tot	528,540	100.00%	528,540	100.00 %

ment used in production. Around 44,000 individuals did not declare any capital goods in 2007, and 43,000 did not declare any in 2008. Moreover, all of the taxpayers in the industry sector declared no capital equipment in both years.

To measure the share of capital equipment in the remaining activities, we calculate the ratio of capital goods to revenues in each year. This ratio is lower than 10% in 2007, but it is much higher in 2008. 1.3 shows that the value of capital goods did not change much between the two years, so most of this change is due to the decrease in revenues starting in 2008, the beginning of the economic crisis.

The combination of the small workforce hired and the low level of capital goods leave no doubt about the true nature of the activities that are included in this database. The database includes only small entrepreneurs who rely almost solely on their workforce, with considerably limited investment in capital goods.

Table 1.3: Average capital invested and ratio capital invested by revenues

	2007		2008	
Economic sector	Capital	Capital/Revenues	Capital	Capital/Revenues
Agriculture	48,240	0.30%	52,529	0.51%
Services	33,470	9.97%	31,354	43.69%
Commerce	36,770	3.85%	37,438	49.51%
Construction	37,442	3.85%	37,454	32.77%
Industry	0	0.00%	0	0.00%
Tot	38,802	8.24%	38,132	30.36%
The table includes data from activities that declared positive values of capital equipment and revenues.				

1.4.2 Panel characteristics

This panel represents the universe of VAT-registered taxpayers who reside for tax purposes in three of the main regions of Italy, Lombardy, Lazio, and Sicily, located in the north, center, and south of Italy, respectively. The periods we observe include tax years 2007 and 2011, for a total of 2,642,700 observations. The panel is perfectly balanced, that is, it includes exclusively taxpayers who issue their tax returns for all of the five years [6].

For each observation (a taxpayer in a given year) we have the following information:

- gender, age, municipality, province and region of residence, economic sector of operation and taxable income type (from self-employment and/or from sole proprietorship);

[6]The lack of any attrition may, at first sight, be seen as a limitation of our dataset, since it prevents us from evaluating the impact on new economic activities and on taxpayers' decisions to interrupt their activities. However, the latter would be empirically difficult to evaluate due to the possibility that the after-audit effect depends on the economic conditions of the business and not on the audit itself. It is noteworthy that DeBacker et al. (2015b) exclude new and ceased corporations from their analysis of the impact of tax audits

- revenues, costs, reported income, tax base, gross tax, and net tax;
- presumptive turnover as calculated by the Italian Revenue Agency (SDS) (for approximately 80% of the panel) and the presence of dependent workers;
- whether the tax declaration was assisted by a tax consultant, distinguished between three types: no consultancy, consultancy with limited responsibility, and consultancy with full responsibility. By limited responsibility, we mean that the tax consultant is responsible only for the correct and on-time presentation of the tax declaration. By full responsibility, we mean that the tax consultant assumes formal responsibility for the correctness of the content of the tax return and is presumably paid an additional fee for bearing this responsibility. Note that, in both of these cases, the tax consultant may have provided some "avoidance services" to the taxpayer.

As we argue in Section 1.7, matching requires selecting variables that simultaneously affect the selection process *and* the outcome variable. For reasons to be described in Section 1.7, we choose the variables listed and described in Table 1.4

Table 1.4: Description of variables

Type of variable	Notation	Description
Taxable income	y	Outcome variable
Gender	Dmale	=1 if male
Age	age	age (in 2007)
Region	DRegion	=1 if resident in Sicily, Lazio, or Lombardy
Sector	DNP	=1 if not a primary sector (not industrial or agricultural)
Marginal tax rate	Mrate	IRPEF marginal tax rate (between 23% and 43%)
Dep.workers	Ddw	=1 if at least one dependent worker
Consultancy	consul	=0 if no consultancy, =1 if limited respons., = if full respons.

Our database contains also information on the occurrence of tax audits conducted on tax returns referring to either base year 2007 or base year 2008. In our dataset, we observe a total of 31,135 audited tax declarations referring to base years 2007 and 2008. Of these, 20,140 are single audits as defined above, and their distributions among audit years and base years are described in Table 1.5[7].

The percentage of single audits ranges from 0.1% to 1.3%. For both base years, it is quite evident that audits are concentrated in audit years that are closer to the expiration deadline (i.e., in audit year 2012 for base year 2007 and in audit year 2013 for base year 2008). Overall, single audits amount to 2.3% and 1.7% of total tax returns issued for the years 2007 and 2008, respectively (excluding in both cases those returns involved in multiple or repeated audits).

[7]Observations of multiple audits as well as those of repeated audits on the same taxpayer are not reported.

Table 1.5: Distribution of single audits

Audit year	Base year	Single audits (number)	Single audits (% of audited+non-audited
2009	2007	452	0.09
2010	2007	1,052	0.209
2010	2008	175	0.03
2011	2007	2,290	0.43
2011	2008	877	0.17
2012	2007	5,139	0.97
2012	2008	3,198	0.61
2013	2008	5,601	1.06

These percentages are in line with the international evidence on the probability of audit. For example, in the U.S., the IRS audits approximately 1% of returns issued by self-employed individuals and sole proprietorships.

Tables 1.6 and 1.7 report the average values for the outcome variable, taxable income (y), and other variables that are described in Table 1.4.

Table 1.6: Descriptions of the outcome variable and covariates, base year 2007

AY2009	**y**	**Dmale**	**age**	**DSIC**	**DLAZ**	**DNP**	**mrate**	**Ddw**	**consul**
Audited	46418.41	0.81	47.26	0.37	0.17	0 .86	0.31	0.39	0.96
All	33915.62	0.77	46.02	0.21	0.25	0 .86	0.30	0.24	0.97
AY2010	**y**	**Dmale**	**age**	**DSIC**	**DLAZ**	**DNP**	**mrate**	**Ddw**	**consul**
Audited	30437.72	0.79	46.84	0.40	0.26	0 .89	0.28	0.42	0.98
All	33890.96	0.77	46.02	0.21	0.25	0 .86	0.30	0.24	0.97
AY2011	**y**	**Dmale**	**age**	**DSIC**	**DLAZ**	**DNP**	**mrate**	**Ddw**	**consul**
Audited	34295.45	0.78	47.14	0.36	0.25	0 .90	0.28	0.24	0.97
All	33903.59	0.77	46.03	0.21	0.25	0 .86	0.30	0.41	0.98
AY2012	**y**	**Dmale**	**age**	**DSIC**	**DLAZ**	**DNP**	**mrate**	**Ddw**	**consul**
Audited	43980.55	0.80	47.55	0.27	0.30	0 .89	0.31	0.24	0.97
All	34033.33	0.77	46.04	0.21	0.26	0 .86	0.30	0.33	0.97

We make two observations.

First, the differences in the outcome variable, taxable income, are hetero-

Table 1.7: Descriptions of the outcome variable and covariates, base year 2008

AY2010	**y**	**Dmale**	**age**	**DSIC**	**DLAZ**	**DNP**	**mrate**	**Ddw**	**consul**
Audited	33927.86	0.84	47.63	0.32	0.22	0.89	0.29	0.40	0.98
All	34156.61	0.77	46.02	0.21	0.25	0 .86	0.30	0.26	0.97

AY2011	**y**	**Dmale**	**age**	**DSIC**	**DLAZ**	**DNP**	**mrate**	**Ddw**	**consul**
Audited	31970.25	0.84	46.15	0.23	0.25	0 .91	0.28	0.41	0.98
All	34139.89	0.77	46.02	0.21	0.25	0 ¿ .86	0.30	0.25	0.97

AY2012	**y**	**Dmale**	**age**	**DSIC**	**DLAZ**	**DNP**	**mrate**	**Ddw**	**consul**
Audited	36377.50	0.81	47.29	0.25	0.22	0 ¿ .90	0.30	0.40	0.97
All	34063.91	0.77	46.01	0.20	0.25	0 ¿ .86	0.30	0.25	0.97

AY2013	**y**	**Dmale**	**age**	**DSIC**	**DLAZ**	**DNP**	**mrate**	**Ddw**	**consul**
Audited	51130.56	0.79	47.93	0.30	0.33	0 .91	0.32	0.39	0.97
All	34342.74	0.77	46.04	0.21	0.26	0 ¿ .87	0.30	0.26	0.97

geneous across audit years. In some audit years, for example, 2013 and 2012, taxable income as reported by audited taxpayers is higher than that reported by non-audited taxpayers. Since these are the years in which the majority of audits are concentrated, this evidence goes against the idea that the revenue agency focuses on taxpayers reporting lower incomes.

Second, the differences in some covariates, namely the region and the presence of dependent workers, are also quite remarkable before matching, thus stressing the need to carefully assess the quality of the matching procedure.

1.5 The challenge of non-random audit data

Audits are usually selected according to two opposite selection rules. Audits can be based on an exogenous (or random) selection rule or on an endogenous selection rule in which specific taxpayer characteristics reported on the taxpayer's submitted return form the basis for audit selection. Usually, both selection rules are implemented in the same time period. However, the presence of such

an endogenous audit selection rule means, of course, that the likelihood of audit selection is an endogenous variable, and this potential endogeneity must be considered in the empirical estimation.

By way of explanation, the audit data used in this study are based on "operational audits" of tax returns selected in a non-random way. Specifically, the audits are determined on the basis of information contained in the tax returns, as opposed to audits that are based on randomly selected tax returns.

Since our audits are non-random, we choose a methodology that does not require randomness. This methodology is summarized by Caliendo and Kopeinig (2008), who propose a combination of matching and differences in differences (DiD) estimators. Treated subjects are first matched with ex ante similar (before the treatment) non-treated subjects. Different matching options can be implemented (radius, nearest neighborhood, stratification, kernel, etc.), and all of them have pros and cons. It is important that in the end there are congruous numbers of treated and untreated matched individuals to investigate the average effect of the treatment on the treated. To deal with the problem of non-random audits, we must implement a procedure that aims to measure the difference between the level of evasion if an audit does or does not occur. However, precisely measuring this difference is impossible because a taxpayer is either audited or unaudited and because we do not observe the true income. To overcome this problem, we use a vector of observables (under the assumption of unconfoundedness) to match audited and unaudited taxpayers before the audit occurs. Then, we deal with the unobservability of the true incomes by assuming that, on average, the difference between the true incomes of audited and unaudited taxpayers is the same after the audit. In this way, the difference between the true and the reported incomes of unaudited and audited taxpayers after the audit will be equal, on average, to the difference between the reported

incomes. Thus, it will then be possible to estimate the difference between the difference in the true and reported incomes of the unaudited taxpayers and the same difference for the audited taxpayers.

This methodology overcomes the problem that one subject can either be treated or non-treated. To effectively estimate the effect of a treatment, ideally both treatment and non-treatment outcomes would be observed for the same subject. Since doing so is clearly not possible, the methodology used in this study first identifies the most similar individuals in the treatment and non-treatment groups and then estimates the net effect of the treatment.

To effectively implement this methodology, it is necessary to have a large group of non-participant individuals who are similar to the participant individuals for some characteristics. These similar characteristics are crucial since they are going to be used to match the two different cohorts of individuals. Moreover, when determining the correct variables to use for matching, they should not be too good. To effectively work, there should be some randomness in the data, that is, some subjects should be able to receive the treatment or not so that at least two individuals with identical characteristics may exist in both states.

Provided that pool of individuals included in the database is appropriate considering the latest observations of Caliendo and Kopeinig (2008), the analysis in this study allows for the investigation of the effect of an audit. The size of the database allows for a congruous number of matches that provides the basis for a rigorous analysis.

An alternative approach is used in another study (Mazzolini et al. (2017)) that is part of the same research project and that uses the same data used here. This alternative approach is a panel approach where difference-in-differences estimation is combined with fixed effects, allowing the authors to control for unobservable time-invariant characteristics of audited taxpayers that may have

caused the Revenue Agency to select them for the audit. On the contrary, in this analysis, we use a year-by-year matching procedure where the matching algorithm includes both time-invariant (gender, sector, and region) and time-variant (age, marginal tax rate, presence of dependent workers, and type of consultancy) characteristics.

In the companion study (Mazzolini et al. (2017)), a semiparametric ex-ante approach restricts the control sample using the nearest-neighbor matching algorithm of Abadie et al. (2004). This empirical strategy enables the authors to address endogeneity related to time-invariant factors influencing the treatment assignment. Specifically, they find a net tax-revenue effect of audits that ranges from three thousand euro (unweighted average) to around eight thousand euro (when weighted by audit outcome).

It is possible to anticipate the result of Mazzolini et al. (2017) in Table 8, which reports the positive effect of audits on the audited years.

1.6 Identification strategy

Our intention is to estimate the impact of an audit conducted at time t on a taxpayer's choice to evade taxes at time $t+s, s = 1, .., n$. We denote the occurrence of an audit (treatment) as $a^t = 1$, the absence of an audit (no treatment) as $a^t = 0$, true income as Y^{TA}, and reported income by the audited taxpayer as Y^{RA}. Ideally, for every audited taxpayer, we would like to measure

$$[(Y^{TA}_{t+s} - Y^{RA}_{t+s})|a^0 = 0] - [(Y^{TA}_{t+s} - Y^{RA}_{t+s})|a^0 = 1] \tag{1.4}$$

where $[(Y^{TA}_{t+s} - Y^{RA}_{t+s})|a^0 = 1]$ is the evasion actually chosen by the audited taxpayer and $[(Y^{TA}_{t+s} - Y^{RA}_{t+s})|a^0 = 0]$ is the evasion that the audited taxpayer would have chosen at time $t+s$ if she had not been audited at time t. The

Table 1.8: Average Audit Effect; Mazzolini et al. (2017)

	OLS		Fixed Effects	
	w/o control	with controls	w/o control	with controls
Audit effect	4,789.22	7,591.91	3,079.46	3,471.90
	[580.166]	[556.275]	[372.334]	[371.373]
Constant	37,401.05	-49,561.44	37,401.28	-44,486.05
	[117.195]	[1,265.478]	[53.583]	[2,502.199]
Ind. controls	No	Yes	No	Yes
Year FE	Yes	Yes	Yes	Yes
Province FE	No	Yes	No	Yes
Industry FE	No	Yes	No	Yes
Individual FE	No	No	Yes	Yes
Observations	2,352,860	2,352,860	2,352,860	2,352,860
R-squared	0.003	0.066	0.013	0.017
Number of id			470,572	470,572

Columns 1 and 2 report OLS estimates. Columns 3 and 4 report fixed-effect estimates. Robust standard errors are in parentheses, clustered at the taxpayer level. Individual controls are gender, age, and age squared.

average treatment effect on the treated would thus be equal to the mean of (1.4) calculated over all audited taxpayers.

However, equation (1.4) cannot be estimated because

1. a taxpayer in a given year is either audited or unaudited, so that $[(Y_{t+s}^{TA} - Y_{t+s}^{RA})|a^0 = 1]$ cannot be observed and
2. true income cannot be observed.

To address issue 1, we denote by Y_{t+s}^{TU} and Y_{t+s}^{RU} true and reported income, respectively, at time $t+s$ for a taxpayer who is *not audited (U)* at time t. We assume that there exists a function $g(X)$, where X is a vector of observables, such that

$$[((Y_{t+s}^{TA} - Y_{t+s}^{RA})|a^0 = 0) = Y_{t+s}^{TU} - Y_{t+s}^{RU}]|g(X_t^A) = g(X_t^U) \tag{1.5}$$

This is the assumption of unconfoundedness; conditional upon having the same $g(X_t)$, the only difference in the amount of evasion chosen at time $t+s$ by an audited and a non-audited taxpayer is due to the treatment, that is, the occurrence of the audit. Let us define as *matched(m)* the non-audited taxpayer who has the same value of X_t as the audited taxpayer that we observe. Thus, we can rewrite (1.4) as

$$[Y_{t+s}^{TU(m)} - Y_{t+s}^{RU(m)}] - [Y_{t+s}^{TA} - Y_{t+s}^{RA}] = [Y_{t+s}^{RA} - Y_{t+s}^{RU(m)}] - [Y_{t+s}^{TA} - Y_{t+s}^{TU(m)}] \quad (1.6)$$

We now need to deal with the issue of the *unobservability of true incomes*. The average value of 1.6 could be estimated if we could assume that

$$E[Y_{t+s}^{TA}] = E[Y_{t+s}^{TU(m)}] \quad (1.7)$$

that is, if, on average, matched unaudited taxpayers have the same true incomes as audited taxpayers. However, this assumption is rather unrealistic, so we revert to estimating the *difference-in-differences* using observations from year t.

Thus, instead of (1.6), we aim to estimate the difference between the change in evasion for unaudited taxpayers

$$[Y_{t+s}^{TU(m)} - Y_{t+s}^{RU(m)}] - [Y_t^{TU(m)} - Y_t^{RU(m)}] \quad (1.8)$$

and the same change evaluated for audited taxpayers

$$[Y_{t+s}^{TA} - Y_{t+s}^{RA}] - [Y_t^{TA} - Y_t^{RA}] \quad (1.9)$$

.[8]

[8]Note that in equations (1.8) and (1.9), we are ignoring that the audits conducted at time

By assuming

$$E[Y_{t+s}^{TA} - Y_t^{TA}] = E[Y_{t+s}^{TU(m)} - Y_t^{TU(m)}] \tag{1.10}$$

that is, that the change in true income for audited and (matched) unaudited taxpayers is the same on average, then the average of the difference between (1.8) and (1.9) can be estimated as the difference in the difference in reported incomes, or

$$E[Y_{t+s}^{RA} - Y_t^{RA}] - E[Y_{t+s}^{RU(m)} - Y_t^{RU(m)}] \tag{1.11}$$

To sum up, our identification strategy is based on a combination of *matching* and of *difference-in-differences*. This approach is known in the literature as *conditional DiD* or as the *DiD matching estimator* (see Caliendo and Kopeinig (2008), p. 55, and the literature cited therein).[9]

1.7 Matching

1.7.1 Estimation of the propensity score

The first step in estimating the propensity score is choosing the underlying model, which includes both choosing a probit or logit specification and, most importantly, choosing the variables to be included. Following the literature Caliendo and Kopeinig (2008), we include, in principle, only variables that should *simultaneously* affect the participation decision and the outcome. In this case, we include variables affecting the probability that a given tax report

t usually concern tax returns issued for years before t. This is the difference between the *audit* and the *base* year that we analyzed in Section 1.4.

[9]Note that this estimator can be implemented either by estimating equation 1.11 or by constructing matching estimates in both periods and calculating the difference between them. The latter procedure is based on the estimation of $E[Y_{t+s}^{RA} - Y_{t+s}^{RU(m)}] - E[Y_t^{RA(m)} - Y_t^{RU(m)}]$, which is equivalent to equation 1.11 if independence is assumed.

is selected to be audited by the Revenue Agency (the participation decision) and evasion as decided by the taxpayer (the outcome).

Although the variables chosen by AE to select taxpayers to be audited are the result of a complex process described in Section 1.3 and, thus, are not fully known, some of them have been recently disclosed (AE, 2015) as follows:

- reported income;
- tax rate;
- region;
- sector;
- congruity to business studies;
- types of consumption;
- specific costs;
- VAT number.

As for the evasion decision, a large literature must to be taken into account. Here, we briefly recapture the main measurable determinants emerging from this literature.

The standard A-S model suggests considering, as determinants of the evasion choice, the perceived probability of being audited, the marginal tax rate, and risk aversion. For Italy, the probability of being audited is likely to be perceived as associated with the congruity status, the *sector of operation*, as well as with the *region* and *province* of residence. The marginal tax rate can be measured, in our database, only with respect to *reported income*, and risk aversion is captured using *age* and *gender*.

As is widely known, in the last twenty years the standard model has been criticized, and explanations of tax evasion based on social interactions have

been put forward (for a summary, see Hashimzade et al. (2014)). This stream of research has led to the consideration of a number of variables as potential determinants of the compliance level, such as:

- avoidance services that can be provided by tax consultants;
- tax morale;
- social customs prevailing in the context where the taxpayer operates;
- perceived fairness of the tax system.

Additionally, a recent literature has emphasized the importance of structural variables, such as the incentives operating through the VAT paper trail (Pomeranz (2013)) for *business-to-business activities* and the presence of *dependent workers* that could trigger whistleblowing risk (Kleven et al., 2009).

In our dataset, we can observe some types of consultancy as well as the presence of dependent workers (see Section 1.4), but we observe at an individual level neither tax morale nor social customs nor the share of B2B transactions.

Thus, we include the variables that simultaneously affect the participation decision and the outcome decision and are measurable. However, we need to perform this selection ensuring that variables are affected by the participation decision or its anticipation (Caliendo and Kopeinig, 2008, p. 38). In particular, congruity status depends partly on a variable (the level of reported turnover) that a taxpayer chooses, and, thus, it is endogenous to the participation decision. A non-congruous taxpayer knows that he has a higher probability of being audited. For this reason, we do not include congruity status. Furthermore, we do not include among the regressors income-related variables, such as specific types of costs and consumptions indicated by AE (2015).

The final list of covariates inserted in the matching vector is provided in Table 1.4.

1.8 Discussion of the empirical approach

The validity of our identification strategy rests on assumptions of unconfoundedness (i.e., equation (1.5)) and of equality in average changes in reported incomes (i.e., equation (1.10)).

The first assumption is common to all matching approaches and is equivalent to assuming that the observed outcome does not depend on unobservables. As noted by Caliendo and Kopeinig (2008) (p. 35), this assumption is strong and needs to be justified by the quality of data at hand.

The assumption we make is that matched audited and non-audited taxpayers would evade to the same degree if they were both audited or both non-audited. Thus, the validity of our assumption rests on our ability to use the data to identify the propensity to evade at an individual level. Moreover, since we use the propensity score as our matching criterion, we need to identify the variables that jointly determine the outcome (i.e., evasion) and the probability of treatment (i.e., the probability of being audited). As we argue in subsection 1.7.1, we believe that we are, in principle, able to identify these variables and, thus, to match, for every observed audit year, taxpayers who have the same propensity to evade and the same ex-ante probability of being audited.

However, our claim needs to be strengthened by performing some robustness checks. We exploit the richness of our database to run some placebo tests. We define as an *audit wave* a combination of an audit year, a base year, and a comparison year.

Considering the number of observations,[10] we have a set of 21 observable audit waves, which is illustrated in Tables 1.9 and 1.14, where we further distinguish between *after-audit, A* and *pre-audit, P* years based on the reasoning proposed at the end of subsection 1.3.1.

[10] We exclude audit waves for which the number of observations is too limited to allow for the estimation of the propensity score.

Table 1.9: Audit waves, base year 2007

Audit years↓ Comparison years⟶	2008	2009	2010	2011
2010	D	A	A	A
2011	P	D	A	A
2012	P	P	D	A
A=after-audit year (in italics when few obs. available); P=pre-audit year; D=dubious year				

Table 1.10: Audit waves, base year 2008

Audit years↓ Comparison years⟶	2009	2010	2011
2011	D	A	A
2012	P	D	A
2013	P	P	D
A=after-audit year ; P=pre-audit year; D=dubious year			

Of the 21 audit waves that can, in practice, be used for estimation, nine can provide a test of the impact of the audit, whereas six can be used as placebo tests. In particular, if the audit increases compliance, then we would expect a positive and significant coefficient on the DiD estimator for "A" years and a non-significant coefficient for the "P" years. Finally, in six cases, according to our reasoning above, the impact of the audit is dubious.

The second assumption is that of equality in the average changes in reported incomes (i.e., equation (1.10)). Note that this assumption is independent of that of unconfoundedness. Indeed, even if random audits were available, it would ideally be necessary to evaluate their impact on evasion rather than on reported incomes. Clearly, there is no difference between these two if true income is supposed to be unchanged. Although this assumption is quite strong, it is very common in most empirical studies concerning tax evasion, where, since evasion is not observable, taxable income is used as a response variable even when random audits are available (see Kleven et al. (2011),).

The main advantage that our approach offers with respect to the panel

approach followed by Mazzolini et al. (2017) is that we can obtain results for each audit wave.

If year fixed effects are irrelevant, one would expect that the impact of an audit conducted at time t on a given base year would depend mainly on the time lag between the after-audit year and the audit year.

On the one hand, the impact should be lower on returns referring to year $t-1$ since, for late audits, such tax returns may be subject to partial ex-post adjustments, as explained in Section 1.3.1.

On the other hand, the impact on returns referring to year $t+s$, with $s>0$, may be lower than that on returns referring to year t if there is a "decay rate" in the updating process (Hashimzade et al., 2014), but we can actually observe a maximum of $s=2$, so this decay process may not be visible in our data.

This approach implies moving across the diagonals of Tables 1.9 and 1.10, considering cells where an "A" is reported, and verifying whether the values of the DiD matching estimator are similar in relative or in absolute terms.

Note that similarity would imply that the response to audits is driven mainly by the time lag between the audits and the after-audit year and not by the specific years that are under consideration. However, the economic crisis may also play a role.

The Great Recession hit the Italian economy particularly hard in 2009, when real GDP decreased by over six percentage points, or 4.9 percentage points in nominal terms. In 2008, the economy had not yet been touched by the crisis, whereas 2010 and 2011 were years of relative economic recovery.

The economic crisis may have influenced the response to audits across different *after-audit years* for a given pair of audit and base years. One would expect that, for example, a different response to an audit in year 2009 than in years 2010 or 2011 simply because year 2009 was deeply affected by the economic

crisis.

1.9 Results

1.9.1 Choosing matching algorithm and assessing matching quality

As discussed by Caliendo and Kopeinig (2008), various matching algorithms are commonly used in the literature. Since there is no a priori understanding of which algorithm is the best, the choice must be guided by the nature of the data at hand and by the estimation strategy. Our database is very large, so that algorithms such as *radius* exceed the computational capacity at our disposal. Thus, we opt for *nearest neighbor matching*. Furthermore, since in our estimation strategy we use the approach suggested by Abadie et al. (2004) with the estimation of *teffects* for the five nearest neighbors, we find it natural to use the same algorithm. Thus, we use the *psmatch2* package with the five-nearest-neighbor matching option.

Regardless of which algorithm is chosen, we must check that the algorithm is able to balance the distribution of the relevant variables (i.e., the components of X) that we shall define as covariates in both the treatment and the control group.

Various methods have been proposed in the literature. The basic idea behind all of these approaches is to compare the situation before and after matching and check if any differences remain after conditioning on the propensity score. If there are differences, matching on the propensity score was not (completely) successful, and remedial measures have to be performed (e.g., including interaction terms in the estimation of the propensity score).

A useful package is the *psmatch2* software in Stata. In particular, this

software allows the relatively easy calculation of the following indicators:

- t-tests for the equality of means in the two samples. Before matching, this is an unweighted regression on the whole sample, and significant coefficients are expected. On the contrary, after matching, the regression is weighted using the matching procedure based on the on-support sample, so no significant differences are expected[11];

- the standardized bias (SB), which, for each covariate, is defined as the difference in the sample means of the treated and matched control subsamples as a percentage of the square root of the average sample variances in both groups. One possible problem with this measure is that it does not provide a clear indication of the success of the matching procedure, even though in most empirical studies an SB below 3% or 5% after matching is seen as sufficient (Caliendo and Kopeinig, 2008, p.48);

- the variance ratio, which is calculated only for continuous covariates and which should equal one if there is perfect balance. An asterisk indicates variables with variance ratios that exceed the 2.5th or 97.5th percentiles of the F-distribution of the relevant test statistic;

- the pseudo R^2 obtained by re-estimating the propensity score on the matched sample (i.e., the sample of participants and matched nonparticipants) and comparing the pseudo R^2s before and after matching. The pseudo R^2 indicates how well the vector of regressors X explains the participation probability. After matching, there should be no systematic differences in the distribution of covariates between both groups, and, thus, the pseudo R^2 should be fairly low;

[11] Note that this is the same test as that performed within the *psmatch* option, but, in that case, the t-test is performed within every block (i.e., every interval in which the observations have comparable values of the estimated propensity score). With our database, however, this operation generates a matrix of hundreds of cells for each audit wave and for every covariate in every block, but no simple criterion for the critical percentage of significant tests is available.

- a likelihood ratio test of the joint significance of all regressors in the probit or logit model. The test should not be rejected before matching but should be rejected after matching.

Table 1.11: Matching quality tests, matched observations

	t-test*		stand.bias**		var.ratio***		pseudo R^2		^likel. ratio	
base year⟶ audit year↓	2007	2008	2007	2008	2007	2008	2007	2008	2007	2008
2009	0	n.a	3%	n.a	0	n.a	0.1%	n.a	1	n.a
2010	0	0	1.1%	1.6%	1	1	0	0	1	1
2011	0	0	2%	2.1%	0	1	0	0	0.999	0.999
2012	0	0	2.8%	2.4%	0	0	0	0	0.948	0.992
2013	n.a	0	n.a	2.4%	n.a	1	n.a	0.000	n.a	0.799
* number of covariates for which p-value is significant at the 5% level										
** highest % std. bias for *matched* obs.										
*** no of continuous covariates for which the variance ratios > the 2.5 and 97.5 pctile of the F dbn. for *unmatched* obs.										
^p-value for the χ^2 statistic for the likelihood ratio test before matching										

Table 1.12: Matching quality tests, unmatched observations

	t-test*		stand.bias**		var.ratio***		pseudo R^2		^likel. ratio	
base year⟶ audit year↓	2007	2008	2007	2008	2007	2008	2007	2008	2007	2008
2009	7	n.a	31%	n.a	0	n.a	2.1%	n.a	0	n.a
2010	7	6	50.1%	39%	0	1	4%	2.8%	0	0
2011	9	9	44%	115,4%	0	2	3,8%	11,5%	0	0
2012	8	10	18.6%	40.2%	2	2	1,7%	3,7%	0	0
2013	n.a	9	n.a	28.7%	n.a	2	n.a	3.0%	n.a	0
* no of covariates with p-value significant at the 5% level										
** highest absolute value of standard. bias for *unmatched* obs.										
*** no of continuous covariates for which the variance ratios > the 2.5 and 97.5 pctile of the F dbn. for *unmatched* obs.										
^p-value for the χ^2 statistic for the likelihood ratio test before matching										

These results are overall in line with expectations, and they confirm that the matching procedure has made it possible to compare the comparables.

1.9.2 Estimates

After matching, estimates are obtained using the new *teffects psmatch* command available in Stata. This command has one very important advantage over the traditional *psmatch2* command. It takes into account the fact that propensity scores are estimated rather than known when calculating standard errors. This difference often turns out to be significant, sometimes in surprising ways. We implement this command using the five best nearest-neighbor matches, as suggested in Abadie et al. (2004).

Tables 15 and 16 report the estimated value of the DiD estimation shown in 1.11 and expressed in euros.

Table 1.13: Estimation of treatment effect, various audit waves, single audits, base year 2007

Comparison years⟶ Audit years ↓	2008	2009	2010	2011
2009	na	na	na	na
2010	1,740 (7,4%)	473 (67.7%)	3,101 (8.7%)	4,242 (3.6%)
2011	541.7 (48.2%)	429.9 (52.8%)	2665.5(1.1%)	3032.0(0.1%)
2012	1335.3 (4.9%)	855.0 (27.9%)	1202.65 (17.1%)	2549.7 (0.6%)
p-values with robust SE in brackets				

Table 1.14: Estimation of treatment effect, various audit waves, single audits, base year 2008

Comparison years⟶ Audit years ↓	2008	2009	2010
2011	119.2 (87%)	2,845.1 (0.1%)	4,448.3 (0.0%)
2012	2,251.8 (24.9%)	1938.0 (10.8%)	2,863.7 (1.1%)
2013	11.06 (98.9%)	1027.8 (17.6%)	2,033.7 (1.3%)
A=after-audit year ; P=pre-audit year; p-values with robust SE in brackets			

Overall, our results are summarized in Table 1.15.

Table 1.15: Summary of results, base years 2007 and 2008

Type of comparison year	Obs.	Sig. at 10%	Non Sig. at 10%
A	9	8	1
P	6	1	5
D	6	2	4
A=after-audit year ; P=pre-audit year			

The hypothesis that audits do change subsequent compliance seems confirmed, since in eight out of nine of the after-audit years, a positive and significant DiD estimate is found, whereas in five out of six placebo years, a non-significant DiD estimate is found. Furthermore, the dubious nature of year $t-2$ seems to be confirmed, since two out of six of these years have a significant DiD estimate, whereas four do not.

1.9.3 Robustness using other matching approaches

There are many possible approaches to matching, and we consider some alternatives.

Regression adjustment (RA) is a methodology that aims to use sample means to estimate the treatment effect. By estimating a regression model adjusted on selected covariates, RA predicts the potential outcomes of a treatment. Specifically, the treated and untreated observations are used to fit linear regressions that predict the outcome of receiving or not receiving the treatment.

RA is a two-step methodology. The first step is the estimation of two different regression models for the outcomes of different covariates of the treated and untreated population. The second step is the estimation of the averages of the predicted for each subject for each treatment level. The difference in these averages provides quantification of the average treatment effect on the treated. A brief description of this methodology is presented in Caliendo and Kopeinig (2008), along with a short literature.

To estimate these results, we use the previously mentioned command available in Stata, *teffects psmatch*. Moreover, we use the default settings in the program and let the software estimate the outcome.

Table 1.16: Estimation of treatment effect, various audit waves, single audits, regression adjustment, base year 2007

	Comparison year							
Audit year	2008		2009		2010		2011	
	value	se	value	se	value	se	value	se
2009	na	na	na	na	na	na	na	na
2010	1,995.3	3.8%	815.2	47.4%	3,477.8	5.3%	4,288.9	3.2%
2011	580.2	45.2%	565.4	41.1%	2,778.1	0.7%	3,043.7	0.1%
2012	1,284.6	5.9%	847.6	28.5%	1,467.0	9.1%	2,724.7	0.2%

Table 1.17: Estimation of treatment effect, various audit waves, single audits, regression adjustment, base year 2008

	Comparison year					
Audit year	2009		2010		2011	
	value	se	value	se	value	se
2011	317.5	64.9%	2,878.6	0.1%	4,393.2	0.0%
2012	2,589.6	18.2%	2,014.9	9.1%	2,852.4	1.1%
2013	96.3	90.4%	1,067.9	16.7%	2,046.0	1.4%

Another estimate is carried out using the Stata command *teffects psmatch* to perform radius matching using the propensity score values used previously. Instead of following the settings in Abadie et al. (2004), the maximum number of possible neighbors is set equal to 100 with a caliper equal to 0.01. Using these settings, it is possible to have a large numbers of matches within a certain radius. Unfortunately, no estimations were feasible for the base year 2008, so Table 20 reports only the estimates for base year 2007.

These results are consistent with the previous ones obtained following the

Table 1.18: Estimation of treatment effect, various audit waves, single audits, radius matching, base year 2007

	Comparison year							
Audit year	2008		2009		2010		2011	
	value	se	value	se	value	se	value	se
2009	na	na	na	na	na	na	na	na
2010	1,751.4	7.6%	361.0	75.6%	2,932.7	10.7%	4,229.5	3.7%
2011	536.7	49.0%	481.7	48.4%	2,824.5	0.7%	3,131.3	0.1%
2012	1,546.8	2.5%	1,197.6	14.6%	1,389.7	11.3%	2,797.8	0.2%

suggestions provided by Abadie et al. (2004). The *teffects psmatch* command allows for other methodologies as well, but they could not be implemented because of the extensive resources required.

These results are similar to those obtained by Mazzolini et al. (2017) who, using a different approach, obtains an overall average estimate of approximately 3.5 thousand euros of additional compliance generated by audits on average in the period between 2007 and 2011.

However, as outlined above, our approach allows us to further explore the nature of the audit impact.

First, it seems that the impact is driven mainly, although not exclusively, by the time lag between the audit year and the after-audit year. More precisely, the impact of an audit conducted in year t can be summarized as:

- an increase in income reported in tax returns referring to year $t-1$ by approximately 2.5-3 thousand euros;
- an higher increase in income reported in tax returns referring to years t and $t+1$, ranging from approximately 3 to approximately 4.5 thousand euros;

Note also that, although the *absolute* DiD values are quite constant for a

given time lag, they change substantially *relative* to average reported values [12].

The second result is that the only after-audit year for which we obtain a non-significant result is 2009, which is the peak of the economic crisis in Italy. Thus, it seems plausible that the Great Recession has limited the impact of audits on subsequent tax compliance.

1.10 Concluding remarks and directions for future research

In this study, we use a unique database to empirically investigate the impact of audits on subsequent tax compliance in Italy. Our empirical approach uses a combination of matching and difference-in-differences techniques and exploits the richness of our database to run some placebo tests.

We show that, on average, the impact of audits on subsequent compliance in Italy between 2007 and 2011 is positive and significant, that it is mainly driven by the time lag between the audit year and the after-audit year, and that it is possibly limited by the economic crisis.

This result is interesting outside the Italian context as well. As we argued before, Italy is a country with low tax morale and low efficiency of tax institutions. Ceteris paribus, we expect operational audits in countries such as the US and Denmark to have greater effects that those estimated here.

This intuition is somewhat confirmed by Beer et al. (2015). They use operational audits conducted by the IRS and provide evidence that these audits have important long-term revenue implications. Three years after an audit, the average small business taxpayer reports around 20 percent more income. This amount is more than twice that estimated by DeBacker et al. (2015a) using

[12] A proper evaluation of the relative impact of the audit would require using logs rather than levels. However, doing so would exclude negative differences.

random rather than operational audits.

A further development of this work could utilize the data to determine the underlying criteria by which tax returns are selected for audit by the Italian Revenue Agency. Since the panel data set includes the universe of taxpayers who reside in three of the main regions of Italy (Lombardy, Lazio, and Sicily) over the tax years 2007-2011 as well as the tax returns selected for audit by the Italian Revenue Agency, it is possible to combine these data to estimate the (observable) factors that determine audit selection. Only a few studies have been able to estimate such audit selection criteria, and such an investigation will certainly enrich this field.

Appendix 1.A The use of true income

The previous section illustrates the empirical strategy that has been implemented following the existing literature to extract the most from the observations included in the database. An alternative approach could be based on true income, as follows:

- Observe taxpayers who are audited in an specific year (say 2007);
- Measure the difference between their reported incomes in 2007 and their true incomes in 2007 (i.e., their incomes as assessed by the Revenue Agency after the audit);
- Observe their reported incomes in subsequent years.

However, true income would be observable only for a minority of audited taxpayers.

Tables 9 and 10 show the audits that take place, the differences between true and declared incomes, and the relative changes in the taxes due. The audited

taxpayers comprise a minority of the 528,540 total taxpayers included in the database, as the table reports.

Table 1.19: Total number of audits, with single audits in brackets

	Base year	
Audit year	2007	2008
2008	69	
	(56)	
2009	630	47
	(452)	(15)
2010	1,551	582
	(1,052)	(175)
2011	3,625	3,039
	(2,290)	(2,073)
2012	8,199	5,416
	(5,139)	(3,198)
2013	161	7,816
	(89)	(5,601)
Total	14,235	16,900
	(9,078)	(11,062)

Table 1.20: Differences between declared and audited incomes and the effects on taxes.

	Declared income	Declared tax	%	Additional income	Additional tax	%
2007						
All taxpayers	33,773	11,240	33.28%			
Audited	44,238	22,611	51.11%	24,177	14,360	59.40%
2008						
All taxpayers	34,043	11,579	34.01%			
Audited	44,433	23,627	53.17%	29,943	17,267	57.67%

The possible changes due to an audit are observed for just a limited amount of the sample. Income in 2007 is audited just 14,235 times, and income in 2008

is audited 16,900 times. Since approximately 4% of the sample is audited, the best way to use this information is to highlight the effect of an audit in the following years.

Thus, we use the reported income, as does all of the literature that observes audits.

A possible example is Kleven et al. (2011), who run a randomization test on a sample of Danish taxpayers to verify the effect of an audit. Specifically, following the existing literature, the work considers treated and untreated taxpayers (who are ex ante identical) to understand the component of tax evasion and the effect of an audit after one year. Looking at the amount of evaded income by self-reporting taxpayers for tax year 2006 (reported in 2007), it emerges that the amount of tax evasion is equal to 41.6%. It then appears that the degree of correction caused by an audit is remarkable in other countries as well in close years when audits are random and performed on self-reporting taxpayers, as in our database.

1.A.1 The empirical approach with the use of the "true" income

Apart from the methodology that is implemented before, it is also feasible to implement a different one. As stated before, it is impossible to observe twice the same subject, in this particular case if she is audited or not. It is possible however to observe the change in the predicted income due to the audit using a matching technique.

A different approach can be based on these steps:

- Predict the evasion of the non audited taxpayers using the predictions of a two steps Heckman simulation.
- In the audited year t, the "true" income is estimated. For the audited

taxpayers, it will be sum of the declared income plus the not declared one resulting from the audit. For the not audited taxpayers, it will be the sum of the declared income plus the predicted evaded one.

- In the next years t+n (with n $\geq$ 0) it is estimated the difference between the declared income in the year t+n and the "true" income in year t.
- With a combination of matching and DID the treated subjects, the audited ones, are matched with ex-ante similar not treated ones, the non audited.

This methodology is tested using the audit year 2007 and, as comparisons, years 2009, 2010, 2011 and 2012. The parameters of the first step (prediction of the treatment) of the Heckman simulation are selected considering internal works of RA such as Braiotta et al. (2015). In this paper there is a complete analysis of the tax evasion of self employees taxpayers in Italy through a propensity score analysis. The research of the best variables to include in the Heckman simulation started from this work. Unfortunately some variables relevant inBraiotta et al. (2015) are not in the database (such as the number of luxury cars and the amount of electricity consumption). Some others are not in that work but provided to be a good predictor (such as sex and percentage of not adherence to sds) .

The parameters included in the second step are instead estimated testing the variables included in the database. After various attempts, the variables that emerged as the most significant are: the declared income, the due tax and the percentage of not adherence to sds.

Results are reported in the table 1.21and they are estimated using the Stata command "hekman" together with the options "twostep" and "first".

Unfortunately, the best simulation provide a poor prediction of the evasion. The Pseudo R2 is just equal to 0.0531. It is worth noting that all the other possible combinations of variables ended up in a even worst results. To observe

Table 1.21: Estimation of the "true" income using the Heckman selection model, year 2007.

ratio_correction	Coef.	Std. Err.	z	P¿z
ratio_correction				
perc_noadeg_sds	0.007493	0.003781	1.98	0.048
redd_imp	-0.0001	2.65E-05	-3.9	0
imp_impst_netta	0.000242	6.27E-05	3.86	0
_cons	1.1312	0.403619	2.8	0.005
select				
1.sex	-0.03276	0.009728	-3.37	0.001
perc_noadeg_sds	-0.00221	0.000134	-16.49	0
redd_imp	-2.79E-06	8.14E-07	-3.43	0.001
imp_impst_netta	6.01E-06	1.90E-06	3.16	0.002
congruo	-0.06767	0.011838	-5.72	0
coerente	0.156239	0.008403	18.59	0
tot_pass	-2.06E-07	1.38E-08	-14.9	0
qta_part_iva	-0.03837	0.005965	-6.43	0
imp_iva_dovuta	-1.46E-06	3.15E-07	-4.65	0
imp_iva_credito	-2.04E-06	5.15E-07	-3.96	0
val_agg_fisc	1.58E-07	5.88E-08	2.69	0.007
num_dipendenti	-0.0182	0.001342	-13.56	0
imp_sps_dipend_smp	-3.61E-06	1.98E-07	-18.21	0
cod_reg				
Lazio	-0.45467	0.009347	-48.64	0
Sicilia	-0.45497	0.010124	-44.94	0
_cons	2.32189	0.017829	130.23	0
mills				
lambda	3.961183	3.125948	1.27	0.205
rho	0.04882			
sigma	81.13833			
Pseudo R2	0.0531			

the difference in the "true" income between audited taxpayers and non audited ones, a comparison between the average declare incomes can be useful.

Table 1.22: Average declared and average "true" incomes of audited and non audited taxpayers, year 2007

	Audited Taxpayers	Non audited taxpayers
Delared income	44,238	33,773
"True" income	65,348	41,195

To match the taxpayers, the chosen variables for the simulation are the same ones of the previous part. The RA methodology is implemented because, although the different methodologies provided similar results, the RA proved to be the most quick one in providing the estimations. The results are reported in the following table

Table 1.23: Differences between declared incomes and "true incomes" , RA methodology, different approach.

	Comparison year							
Audit year	2008		2009		2010		2011	
	value	se	value	se	value	se	value	se
2009	na	na	na	na	na	na	na	na
2010	-24,591	0.00%	-25,859	0.00%	-23,060	0.00%	-22,226	0.00%
2011	-21,135	0.00%	-21,235	0.00%	-18,904	0.00%	-18,612	0.00%
2012	-16,462	0.00%	-16,974	0.00%	-16,296	0.00%	-15,026	0.00%

Results proved to be highly influenced by the prediction of the evasion for the non audited taxpayers. As reported before, the predicted correction in the income of the non audited taxpayers is very limited if compared to the one of the audited ones. The DID between the matched individuals then is negatively influenced by the fact that the "true" income of non audited taxpayers is far lower than the "true" income of the audited ones.

Appendix 1.B Complete results for the other matching approaches

2007 Regression adjustment						
Audit 2010	Coef.	R. Std. Err.	z	P - value	[95% Conf	Interval]
diff_08	1,995.3	963.2	2.1	3.80%	107.5	3,883.0
diff_09	815.2	1,138.2	0.7	47.40%	-1,415.7	3,046.1
diff_10	3,477.8	1,795.1	1.9	5.30%	-40.5	6,996.1
diff_11	4,288.9	2,000.8	2.1	3.20%	367.3	8,210.4
Audit 2011	Coef.	R. Std. Err.	z	P - value	[95% Conf	Interval]
diff_08	580.2	771.7	0.8	45.20%	-932.3	2,092.7
diff_09	565.4	687.6	0.8	41.10%	-782.4	1,913.1
diff_10	2,778.1	1,032.3	2.7	0.70%	754.9	4,801.3
diff_11	3,043.7	929.6	3.3	0.10%	1,221.7	4,865.7
Audit 2012	Coef.	R. Std. Err.	z	P - value	[95% Conf	Interval]
diff_08	1,284.6	680.5	1.9	5.90%	-49.2	2,618.4
diff_09	847.6	793.1	1.1	28.50%	-706.9	2,402.1
diff_10	1,467.0	869.3	1.7	9.10%	-236.8	3,170.8
diff_11	2,724.7	894.6	3.1	0.20%	971.3	4,478.1

2008 Regression adjustment						
Audit 2011	Coef.	R. Std. Err.	z	P - value	[95% Conf	Interval]
diff_09	317.5	698.4	0.5	64.90%	-1,051.2	1,686.3
diff_10	2,878.6	838.0	3.4	0.10%	1,236.2	4,521.0
diff_11	4,393.2	919.6	4.8	0.00%	2,590.8	6,195.6
Audit 2012	Coef.	R. Std. Err.	z	P - value	[95% Conf	Interval]
diff_09	2,589.6	1,940.9	1.3	18.20%	-1,214.5	6,393.7
diff_10	2,014.9	1,194.0	1.7	9.10%	-325.2	4,355.0
diff_11	2,852.4	1,116.4	2.6	1.10%	664.4	5,040.5
Audit 2013	Coef.	R. Std. Err.	z	P - value	[95% Conf	Interval]
diff_09	96.3	796.9	0.1	90.40%	-1,465.6	1,658.3
diff_10	1,067.9	771.9	1.4	16.70%	-445.0	2,580.9
diff_11	2,046.0	835.8	2.5	1.40%	407.7	3,684.2

2007 Radius mimic						
Audit 2010	Coef.	R. Std. Err.	z	P - value	[95% Conf	Interval]
diff_08	1,705.7	980.4	1.7	8.20%	-215.8	3,627.3
diff_09	466.5	1,132.9	0.4	68.10%	-1,754.0	2,687.0
diff_10	3,197.1	1,793.8	1.8	7.50%	-318.8	6,712.9
diff_11	4,105.2	2,006.3	2.1	4.10%	173.0	8,037.4
Audit 2011	Coef.	R. Std. Err.	z	P - value	[95% Conf	Interval]
diff_08	320.4	778.4	0.4	68.10%	-1,205.2	1,846.0
diff_09	57.3	696.8	0.1	93.40%	-1,308.4	1,423.1
diff_10	2,371.6	1,045.5	2.3	2.30%	322.5	4,420.7
diff_11	2,629.6	941.9	2.8	0.50%	783.6	4,475.7
Audit 2012	Coef.	R. Std. Err.	z	P - value	[95% Conf	Interval]
diff_08	1,196.3	681.1	1.8	7.90%	-138.7	2,531.2
diff_09	649.4	790.1	0.8	41.10%	-899.3	2,198.0
diff_10	1,371.1	869.4	1.6	11.50%	-332.9	3,075.1
diff_11	2,701.9	890.4	3.0	0.20%	956.8	

Chapter 2

A Minimum Tax in Italy: Impact on Revenues and Redistribution

Urpis, E.

2.1 Introduction

This study investigates the effects of implementing a presumptive tax in a country like Italy, which has a low level of tax compliance compared to other developed countries. The main aim is to study the effect of a change in the policy from a particular starting condition. More specifically, this analysis compares the initial level of taxes collected in Italy from a particular group of taxpayers, to levels that would be collected if the country implements a particular presumptive tax.

This work will then show in which particular case it is convenient to im-

plement presumptive taxation to collect taxes. Given that Italy has some difficulties in collecting taxes due, this study investigates whether implementing a presumptive tax will be a good solution to this problem. Moreover, the presumptive tax proposed here is of a particular form: a minimum tax. The work I present will study the impact of a presumptive tax in the Italian context, and show the viability of a minimum tax implementation and describe the particular case in which it is a viable policy to increase the total amount of taxes collected, while also limiting the drawbacks.

Presumptive taxes are characterized by a change in the tax base. In previous years, they were considered a viable solution to tackle compliance issues in different states. If a tax base is difficult to observe, it is possible to approximate it to one that is easier to measure. However, this shift can bring some drawbacks, the pros and cons of which have been analyzed in recent years; among all Yitzhaki (2007), Slemrod and Yitzhaki (1994) and Balestrino and Galmarini (2005). They can lower the administrative costs for tax agencies and the compliance costs for tax payers; moreover, they are relatively easy to implement. However, at the same time, presumptive taxes lower the degree of precision and cause concerns about both horizontal and vertical equity.

The existing literature focuses mostly on describing presumptive taxes from a theoretical point of view or describing different presumptive tax implementations in the countries that decided to adopt them. Theoretical studies have extensively described the main characteristics of these taxes together with their costs and benefits. Some works illustrate the technical issues of their implementation, while others focus on the impact on specific aspects (such as the effects on labor, influence on business choices, etc.). Moreover, the economic literature provides several contributions illustrating the implementation of presumptive taxes in different contexts.

This work enriches the existing literature with some insight on the effects of presumptive taxes when they are implemented as minimum taxes. More specifically, it provides precise and effective estimates of the implementation by using a large and reliable database. Since implementing a minimum tax would affect the taxpayer behavior, this work will also compute the effects that possible taxpayer reactions might have on the final outcome. This study computes three levels of reactions and illustrates the changes in the results for each.

Using a database provided by the Italian Revenue Agency (RA), this study estimates the impact of a minimum tax on the total amount of taxes collected and calculates the redistributive issues that this would bring. To do so, it implements two different presumptive taxes related to the methodologies developed by Istat and the RA. The analysis concerns a very specific group of taxpayers whose characteristics are discussed in detail.

The work is organized as follows: Section 2 illustrates the theoretical background of this research and defines presumptive taxes and minimum taxes; Section 3 provides a brief description of the two methodologies to estimate the presumptive taxes; Section 4 describes the database, and finally, Section 5 illustrates and discusses the results.

2.2 Literature review

Governments levy presumptive taxes when it is difficult for them to obtain reliable information on the declared tax base, or when the information they collect is not reliable. To address this problem, the taxpayer declared tax base is substituted with a presumptive one, estimated by another body like the public administration, which is easier and more reliable to observe.

The existing literature on presumptive taxation can be divided into three groups: studies that provide a theoretical analysis of the characteristics of pre-

sumptive taxes and the effects on subjects, works that investigate the drivers of the adoption of these taxes, and other studies that describe effective implementations in different contexts.

Among the research in the first group, Yitzhaki (2007) provides a very general and complete definition of presumptive taxes, stating that we have presumptive taxes "... whenever the legislator is using one tax base in order to approximate another." This shift between the tax base and real tax is the core of presumptive taxation. When one activity is difficult to observe or it is possible for the taxpayer to easily a hide part or activity, the government prefers to link the tax of one particular base to one that is easier to observe. Presumptive taxes arise whenever it is hard for a tax agency to collect the proper information or it finds it difficult to implement the tax laws.

Yitzhaki (2007) also provides a list of the pros and cons of presumptive taxation. Presumptive taxes are characterized by simplicity, low compliance costs, and low administrative costs, while still guaranteeing equal opportunity for taxpayers to organize their activity. However, there are also some drawbacks, such as, a lower degree of accuracy and the need constantly update the parameters to estimate the presumptive tax. Some other characteristics are more subtle; for example, there is no final word about horizontal equity. Moreover, when analyzing these taxes, we should consider that it is sometimes difficult to classify a tax presumptive or not since "... all taxes are presumptive to some degree" Slemrod and Yitzhaki (1994).

Other works focus on more specific characteristics of presumptive taxes and past implementations over time. Bulutoglu (1995) discusses the advantages and disadvantages in the context of developing countries. He investigates how a presumptive income is implemented and how presumptive tax models approximate revenues, costs, or other assets. In this work, presumptive taxes are seen

as an alternative to audits, distinguishing when it is not negotiable and when taxpayers can challenge the presumptive tax, who bear the cost of the proof.

Instead, Balestrino and Galmarini (2005) put more emphasis on the behavior of the subject involved. They use an occupational choice model to investigate the effects of implementing a presumptive tax on entrepreneurs and workers. In more detail, they provide an analysis of the redistribution issues and efficiency changes following the negative effects that presumptive taxes have on the resources invested in counteracting tax avoidance.

Some other studies, such as Akerlof (1978), Immonen et al. (1998) and Boadway and Pestieau (2006), focus on more technical details, such as the calibration of the tax and the identification of the groups involved. Among these, Immonen et al. (1998) focuses the analysis on tagging and means testing, investigating the right balance between them.

The second group of studies describe the reasons that might drive the government to implement such taxes. Larin and Jacques (1994) illustrate the three reasons that drive a government's decision to implement these taxes:

- Tax preferences restrictions: an attempt to counter taxpayer planning with counter-planning;
- Attempts at redistribution: a state can attempt to redistribute a minimum amount of money away from taxpayers who have successfully exploited the tax system;
- Administrative countermeasure: as an aid for states with insufficient auditing facilities to investigate taxpayers' strategic tax-planning.

This list makes it possible to understand the reasons that push different states to adopt presumptive taxes. For developing countries, they can be a good way to deal with the lack of resources to invest in audit activities. On the

contrary, developed countries can effectively implement them to let the highest taxpayers pay a minimum amount of taxes. In the latter context, presumptive taxes avoid large investments in tax planning from undermining the progressiveness of a tax system. To support their work, they provide a complete list and the main characteristics of the different presumptive taxes in Europe, Africa, and North and South America, showing how presumptive taxes differ from state to state.

The final group of studies instead describes how presumptive taxes were implemented in different countries. Some, such as Stotsky (1995), illustrates the pros and cons considering various experiences over time. Others are more specific and illustrate the experiences of single countries, such as the study by Gamarra Rondinel (2017), who uses the bunching approach to investigate the responses of the Argentinian firms to the implementation of a simplified regime, Pashev (2006), who describe the experience of Bulgaria, or Wallace et al. (2002), who focuses on Russia.

The object of this work is to study whether introducing a minimum tax can be a viable solution for a country affected by a low degree of tax compliance, such as Italy. The originality of this work is in the econometric analysis that it provides. Some previous studies analyze a database to study the implementation of presumptive taxes. Works such as Larin and Jacques (1994) investigate the effects of a presumptive tax implementation in Canadian presumptive on different groups of taxpayers. Logue and Vettori (2011) investigate the possibility of narrowing the tax gap in the United State by implementing a presumptive tax. This work however will go further by providing an empirical estimation of the effect of the implementation while also considering taxpayers' reactions.

In more detail, this study investigates whether such an introduction will be effective in collecting supposedly undeclared taxes. The choice of implementing

a minimum tax should also account for the costs associated with such a decision. As Slemrod (2007) points out, each tax implementation brings issues of fairness and efficiency. Together with the benefits associated with increases in revenues, this work also considers that the cost of the decreased redistributive effect of taxation causes a loss of disproportionality. The trade-off between an increase in revenues and a decrease in the differences in the shares of taxes is expected.

Finally, it should be remarked that this work adopts two different methodologies to estimate presumptive taxes, which are very specific to the Italian context since they are implemented by two of the largest bodies that estimate the shadow economy. The Istat one is based on Franz (1985)'s pioneering work and aims to estimate the Italian GDP. The methodology developed by Istat to estimate shadow income is illustrated in detail with some estimations by Puggioni (2015). Santoro and Fiorio (2011) and Dal Prato (2016), among others, describe the RA's methodology and the historical background that led Italy to develop this particular methodology to estimate taxpayers' incomes a priori, and illustrate the parameters to estimate the presumptive tax base in detail, with a remarkable reference to the Italian context.

2.3 Theoretical background

2.3.1 Implementation of a minimum tax

The introduction of presumptive taxes brings different effects that require investigation. First, taxpayers can react to the introduction of presumptive taxes; second, a redistributive effect can arise among taxpayers.

To better analyze taxpayers' behavior to the introduction of presumptive taxes, imagine a finite cohort of taxpayers j (with j = 1, 2,. . . ,..n), and denote declared taxes with T_D^j and presumptive taxes with T_P^j. In turn, these can be

written as

$$T_D^j = \tau(y_D^j) y_D^j \tag{2.1}$$

$$T_P^j = \tau(y_P^j) y_P^j \tag{2.2}$$

where $\tau(y^j)$ is the average effective tax rate, y_D^j is the declared tax base, and y_P^j the presumptive tax base. Clearly, if the tax schedule is progressive, $\tau'(.) > 0$, and an increase in the tax base will cause an increase in $\tau(y^j)$. Suppose that the tax rule dictates that, for each j, the tax to be paid is equal to

$$T^j = max(T_D^j, T_P^j) \tag{2.3}$$

This means that the presumptive tax is actually implemented as a minimum tax. Our purpose is to evaluate the impact of introducing this minimum tax instead of a rule that simply prescribes that taxpayers must pay T_P^j. To do so, we consider two changes:

- the change in aggregate tax revenues, ΔR;
- the change in the coefficient of concentration of taxes paid with respect to the ranking of declared incomes, ΔC.

We must conduct the evaluation by observing incomes as declared only before the minimum tax rule is implemented. If taxpayers comply with the new policy and pay the taxes on the maximum between the declared and presumptive taxes, by definition, aggregate tax revenues will increase after the tax reform, that is, we will observe

$$\Delta R = |RP - RD| > 0, \tag{2.4}$$

where RP is the revenue with the minimum tax and RD the declared revenue. To evaluate the change in the concentration of taxes, consider that according to Yitzhaki (2007), the "distributional characteristics of the presumptive tax tend to be worse than those of the regular tax," since "the presumptive tax rate tends to be lower on [...] higher income taxpayers." To capture this inequity in the presumptive tax, we measure the change in the disproportionality effect with respect to declared incomes. For every progressive tax system, Kakwani (1977)'s theorem states that the disproportionality effect can be measured by the difference

$$G^T - G^Y, \tag{2.5}$$

where G^T is the Gini index of taxes paid and G^Y is the Gini index of income declared. Indeed, if the tax and income rankings are the same, a progressive tax system would generate a positive value of (5) because the proportion of taxes paid by any x% of the poorest taxpayers should not exceed the proportion of income declared by the same x% of the poorest taxpayers. In graphical terms, the Lorenz curve of taxes should be further from the 45◦ line than the Lorenz curve of income, thus generating a higher Gini index. In contrast, a flat tax would generate $G^T = G^Y$. This also implies that, for a given ranking of declared incomes, the decrease in the concentration coefficient of taxes denotes a loss in the disproportionality effect of the tax system. Now, in general, since the presumptive taxes tend to be higher on lower incomes, one would expect to observe

$$\Delta C = (C^P - G^Y) - (C^D - G^Y) = C^P - C^D < 0, \tag{2.6}$$

where C^P and C^D are the concentration coefficient [1] of presumptive and

[1] The tax implementation changes the order of taxpayers compared to their earnings. Since

reported taxes, respectively, with respect to the ranking of reported incomes. For these reasons, one expects the minimum tax to decrease the disproportionality of taxes, and, therefore, their ability to redistribute income. The latter statement is debatable if one considers that declared incomes are not true incomes. Thus, the loss in disproportionality may be irrelevant since the benchmark (i.e., declared taxes) is biased by evasion. Nevertheless, declared incomes (and taxes) are a benchmark or reference point that taxpayers use to evaluate the (in)equality of taxes. Following this line of reasoning, the difference $C^P - C^D$ can be interpreted as a measure of the departure from this benchmark.

2.3.2 The equity-efficiency trade-off

As illustrated in the previous subsection, one expects the tax reform to increase revenues but to decrease disproportionality. While the former can be seen as an advantage of the reform (abstracting away from considerations about the efficiency of public spending relative to the decrease in private consumption), the latter can be seen as an equity cost of the reform. As 2.4 shows, implementing a minimum tax increases the total amount of tax collected. This happens because taxpayers with $T_P^j > T_D^j$ pay more taxes, while taxpayers with $T_P^j < T_D^j$ are not. Since the upward correction of taxes hits only some taxpayers, the proportion of taxes paid changes. As 2.6 shows, implementing a minimum tax causes a change in the concentration of taxes paid. The correction reduces the disparity because just some taxpayers pay more taxes; consequently, taxes paid are shared more equally. From 2.4 and 2.6, the trade-off is always present. Moreover, an increase of the amount of tax collected brings a stronger decrease in the concentration.

As stated before, we evaluate this trade-off for the Italian case. To do so,

the Gini index is computed by ranking the values from the lowest to the highest, a comparison of the Gini index of income and taxes will be misleading due to the different rankings. For this reason, it makes more sense to estimate the concentration coefficient of the minimum taxes ordered by declared incomes and then compare it to the Gini index of declared taxes.

however, we must consider the possibility of taxpayers' reactions.

2.3.3 Taxpayers' reactions

To properly evaluate taxpayers' reactions, we should run a natural experiment, comparing reported incomes before and after the introduction of a minimum tax. Unfortunately, our database does not allow for this possibility. Thus, to evaluate the impact of the reform on revenues and the concentration of taxes paid, we require some simplified hypotheses. Namely, two possible reactions are considered:

- A change in the intensive margin: taxpayers with reported incomes above the presumptive incomes would consider paying less taxes by paying the latter (levelling effect).
- A change in the extensive margin: taxpayers with reported incomes below the presumptive ones may find it too expensive to comply with the taxes calculated over the latter incomes and may thus consider leaving the market (leaving effect).

Each taxpayer will react with one of four different behaviors. This study considers a parameter σ inversely related to taxpayers' reactions. This parameter ranges from 0 to 1, with a σ equal to 1 representing no reaction, while a σ equal to 0 represents the maximum reaction. Graphically, for a given value of σ, we can divide the population of taxpayers into five groups:

- Group 1: taxpayers with $T_D^j < T_P^j$ and $(1 - (T_D^j/T_P^j)) > \sigma$. For this group, it is too costly to pay the taxes on the presumptive income and they consequently decide to leave the market [2], so they are affected by the leaving effect.

[2]This study assumes that if the difference between T_D^j and T_P^j is too broad, the taxpayer prefers not to produce goods or services and leaves the market. Since T_D^j are taxes that the

- Group 2: includes taxpayers with $T_D^j < T_P^j$, but who are also characterized by $(1 - (T_D^j/T_P^j)) < \sigma$. These subjects have declared taxes lower than the presumptive ones but decide to stay in business and pay a higher level of taxes.

- Group 3: taxpayers with $T_D^j > T_P^j$ and $(1 - (T_P^j/T_D^j)) < \sigma$. Taxpayers whose declared taxes are higher than the presumptive ones, but who pay the declared taxes since they determine that the cost of a possible audit will be higher than the difference between the two taxes.

- Group 4: taxpayers with $T_D^j > T_P^j$ and $(1 - (T_P^j/T_D^j)) > \sigma$. This group is characterized by higher declared taxes than presumptive taxes, but they pay the presumptive tax since the difference is so high that they find it more profitable to pay the presumptive tax and risk incurring the cost associated with an audit, so they are affected by the levelling effect.

- Obviously, in this setting, subjects with $T_D^j = T_P^j$ simply pay the reported taxes, which are exactly equal to the presumptive taxes.

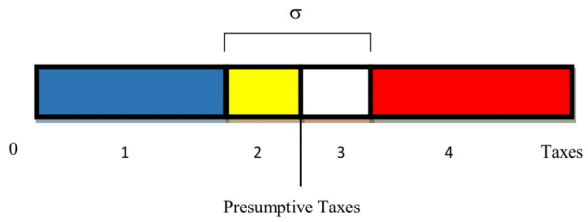


Figure 2.1: Potential distribution of taxpayers

It is possible to represent graphically such dynamics. In fig.2.1the taxes (presumptive and declared) are set in the horizontal axe. We can observe how

taxpayer would declare before the implementation of a minimum tax, while T_P^j are the taxes estimated with a reasonable methodology, the difference between T_D^j and T_P^j is the estimated evasion. The parameter σ then represents taxpayers' willingness with $T_D^j < T_P^j$ to accept the upward correction of taxes.

the magnitude of σ influences the taxpayers' participation in each group.

Looking at the figure, it is possible to see that groups 1 and 4 increase as σ decreases. This happens because the more the taxpayers are reactive, the more they will tend to leave the market or to pay the presumptive taxes because the declared taxes are too high compared in comparison.

These reactions then cause a change in the total amount of income, and their average value changes. Implementing a minimum tax will then cause an income effect on taxpayers. If taxpayers in group 1 leave the market and earn nothing, they will experience the most severe consequences of this implementation. Instead, group 2 will pay more taxes and will thus be affected by a negative income effect. Group 3 will be unaffected, while group 4, paying less taxes, will have a positive income effect. The change in the concentration of the taxes is measured with the concentration coefficient with regard to the ranking of (observed) reported incomes. Finally, we should point out that the collected taxes might change, but not the presumptive taxes, since they are based on observed incomes.

We consider three possible values of the parameter σ: high (85%), medium (50%), and low (15%).

It is also possible to represent the relationship between the declared and presumptive taxes, which appear on the horizontal and vertical axes, respectively, in Figure 2. Taxpayers with $T_D^j = T_P^j$ will be in the bisector, while groups 1 and 2 will be on the left of the bisector and groups 3 and 4 on the right.

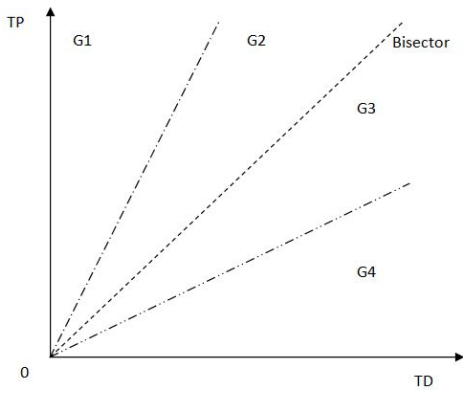


Figure 2.2: Relationship T_D^j and ^j_P

The magnitude of σ then will split groups 1 and 2 and groups 3 and 4. Considering the part of the graph to the left of the bisector, as σ increases, the line dividing the two groups will move toward the y axis. When σ approaches 1, the numerator of $(T_P^j - T_D^j)/T_P^j$ approaches T_P^j. This measures the increase in taxpayers' willingness to stay in the market if they have to pay presumptive taxes higher than the declared taxes. At the same time, an increase in σ causes a shift downward in the line dividing the part of the graph to the right of the bisector. In this case, with σ approaching 1, the numerator of $(T_D^j - T_P^j)/T_D^j)$ approaches T_D^j. This captures the increase in taxpayers' willingness to comply with the policy and pay the declared taxes, even if they are higher than the presumptive ones.

To summarize, the analysis first investigates the implementation of a minimum tax without reactions. The study considers the effects of the change in the total amount of tax collected and the change in the disproportionality. Later, we address the research question in three scenarios: low, medium, and high, corresponding to strong, intermediate, and mild taxpayer reactions.

2.4 Methods

Two different estimations of presumptive taxes are implemented to analyze whether a particular estimator affects the final result. To do this, a corrected income is first computed using two different methodologies, and then the presumptive taxes are calculated for each. It should be remarked that all taxes are estimated following the prescriptions of Italy's income tax. These methodologies were developed by two prestigious Italian associations: Istat and the RA, which are analyzed in detail in the appendix 2.B.

Both methodologies are micro-founded (based on taxpayers' declarations), but their aims differ. Istat developed its methodology to quantify GDP, while RA aims to estimate the amount of taxes that each taxpayer does not declare, since the Istat methodology has a macro purpose (treating the taxpayers in bulk) versus the RA's micro purpose (a personal presumptive tax base is estimated for each taxpayer).

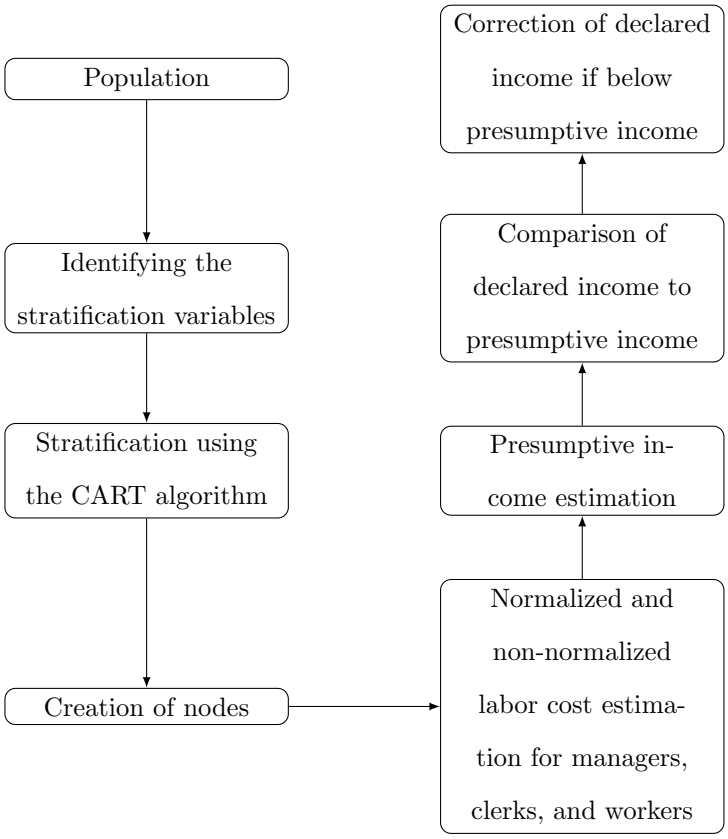
Istat developed a methodology that aims to calculate a threshold level, the shadow income, or the presumptive statistical income, to compare with the declared income and estimate underreported incomes. The main intuition behind this methodology are that entrepreneurs work full time for their firms, which exist in a context of perfect competition. The key assumption is that in normal conditions, the income earned by an entrepreneur or a professional cannot be lower than the income they could earn by working as a dependent worker doing similar tasks. If that happens, then the reported income can be considered to be affected by underreporting.

Following this approach, the income declared should be corrected and increased to at least the salary of a dependent worker. The whole methodology can be considered as a comparison between the declared income and a threshold level, the presumptive statistical income, to estimate underreported amounts.

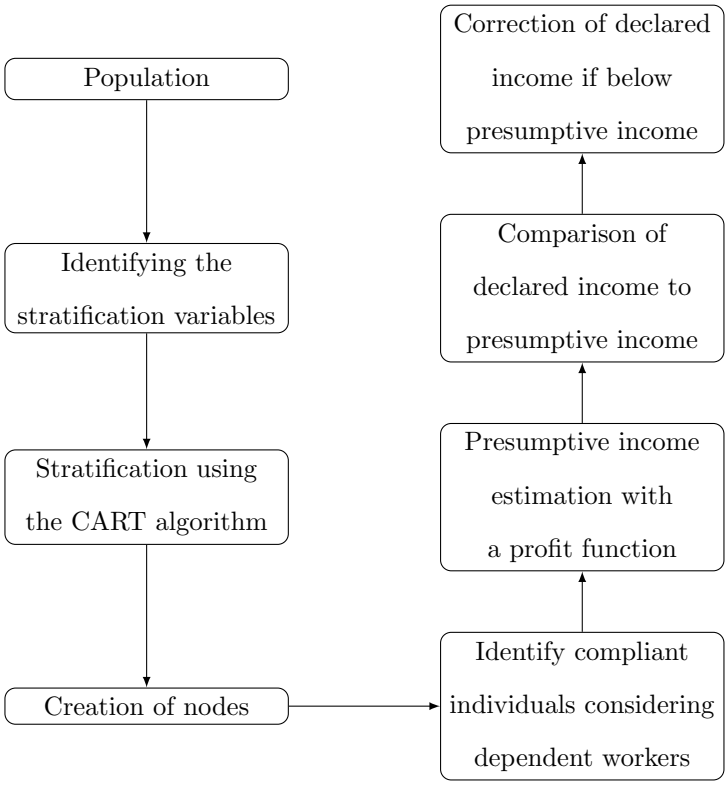
The only requirement is the existence of workers doing similar tasks who can provide a benchmark for this comparison. Once this condition is fulfilled, it is possible to compare the declared income of all entrepreneurs to the presumptive statistical income and correct the supposed underreported declarations upwards to the presumptive income level.

Details of this methodology are included in the appendix 2.B. However, it is possible to briefly illustrate Istat's methodology to estimate the shadow income of taxpayers with virtually not dependent workers when they do and do not have high skill levels and Istat call it G1.

In case the taxpayer has a low skill level, the Istat methodology aims to cluster taxpayers in homogeneous groups to collect the normalized and non-normalized salaries of managers, clerks, and workers to estimate the presumptive income of each cluster. If the declared income of a taxpayer is lower than the presumptive income of the cluster, then the income is corrected upward to the presumptive income.



For taxpayers with high skill levels, the methodology is very similar to the previous one. The main difference is that the information about managers, clerks, and workers salaries are used to identify compliant taxpayers. After identifying compliant taxpayers, it is possible to use the information included in their declaration to set a profit function to estimate presumptive income. Again, after estimating presumptive income, if the declared income is less than the presumptive income of the cluster, then the income is corrected upward to the presumptive income.



The RA methodology differs because it is based on identifying several parameters to estimate the presumptive tax base. The methodology is called the studi di settore (SDS). It first collects various information on internal (costs, structure, productive process, etc.) and external variables (degree of market competition, prices, etc.) of a selected sample of taxpayers. A cluster analysis of this information estimates the presumptive income of each taxpayer by adopting multiple regression techniques. It is important to stress that this procedure is not based on a sample analysis, but each step includes constant research on the most sensitive variables that influence the final result.

Both methodologies aim to correct the declared income of taxpayers who are supposedly non-compliant. However, as the description suggests, the correction might differ since the assumptions and implementation differ. The Istat

methodology corrects the taxpayer clustering by assigning them to homogeneous groups such that taxpayers with below-average incomes are corrected. Moreover, if taxpayers declare an income close to zero, the Istat correction might be severe. The RA methodology instead estimates the presumptive income for each taxpayer considering their personal characteristics. The RA's methodology can correct the information about more taxpayers, that is, it also addresses high-income taxpayers if their incomes are higher than average. Moreover, it can also correct information for very low-income taxpayers less severely compared to the Istat methodology. Using the RA methodology with certain parameters, it is possible to estimate a very limited presumptive income using taxpayers' characteristics. Given their differences, it is expected that the Istat methodology will correct fewer taxpayers, but do so more intensively, while the RA method will correct more taxpayers, but for smaller amounts.

2.5 Data

2.5.1 The database

This study uses a database provided by the RA, Agenzia delle Entrate. This database is a balanced panel that reports various information and statistics declared by Italian residents living in three regions of Italy. In this panel, taxpayers are observed over a period between 2007 and 2011. Moreover, there are no employed subjects: all records refer to self-employed taxpayers and sole proprietorships.

The database is organized in terms of the Istat's methodology to estimate GDP and the real size and characteristics of the Italian economy. Since this study investigates the effects of introducing a minimum tax related to a very specific group, taxpayers will be rearranged in terms of the Istat's methodology

to analyze this group.

The database contains all of the information declared by taxpayers in the different reports they must file throughout the fiscal year. The information concerns demographic (age, sex, etc.) and economic characteristics (returns, workforce, etc.). Moreover, there is also information specific to each taxpayer on prior undeclared income estimated by the RA; the SDS provides an initial insight into the loyalty of a taxpayer.

The dataset consists of a balanced panel of 528,540 selected taxpayers living in Lombardy, Lazio, and Sicily during 2007 and 2011. The taxpayers included in the analysis are those outside the agricultural or related sectors, since revenues in these sectors are estimated on a forfeit basis. Unfortunately, this study cannot use all of these subjects since some of them do not fulfil all the minimum requirements. For example, for 2007, some units were deleted since they had more than 100 dependent workers (37,488) [3], others because they do not provide any information about the SDS (57,041), and so on. After filtering, 434,011 subjects remain for 2007 and the following estimations are based on these units.

This analysis does not include 2008 since it is not possible to calculate the presumptive tax for this year with the Istat methodology.

Following the Istat methodology, it is possible to classify subjects into five groups. All taxpayers must fulfil the Istat specification precisely; any taxpayer missing just one characteristic are placed in Group 0. The appendix describes characteristics of this classification. Because subjects can change characteristics over time, Table 1 provides the classification for 2007, which will be considered as the base year.

[3]Since this work investigates taxpayers with few dimensions, it includes only taxpayers with fewer than 100 dependent workers consistently with the Istat methodology implemented to create the Emens database.

Table 2.1: Distribution of subjects among classifications, 2007.

Group	**Lombardy**	**Lazio**	**Sicily**	**Total**
0	177,466	78,502	57,596	313,564
G1	12,332	6,666	4,623	23,621
G1A	110	66	27	203
G1B	10,780	5,769	4,130	20,679
G1C	1,442	831	466	2,739
G2	30,600	13,792	12,882	57,274
G3	2,809	904	983	4,696
G5	19,215	8,851	5,912	33,978
G5A	1	3	1	5
G5C	17,232	8,120	5,659	31,011
G5D	1,982	728	252	2,962
Total	242,422	108,715	81,996	433,133

A large part of the sample falls into group 0 because these subjects do not fulfil at least one of the requirements for inclusion in the in Istat's classifications, for example, they are too young, do not report all information, there is no indication of congruity, and so on. Moreover, the G4 does not show any units since this group's characteristics (such as being part of an Italian group) are not observed in the database. Additionally, G5B, firms in a particular situation, refers to firms with characteristics that are not included in the database, and thus none of these units appear in the table.

Not all regions of Italy are included; the database covers only three regions Italy: Lombardy, Lazio, and Sicily. This choice was made for two reasons. First, these are the most populated regions of each Italian macro area (North, Center, and South). The three Italian macro areas have homogenous socioeconomic variables within them and heterogeneous characteristics between them. It was

considered suitable to choose representative regions in terms of population to highlight any possible differences among taxpayers related to geographical areas. There are 242,422 subjects from Lombardy, 108,715 from Lazio and 81,996 from Sicily. Taxpayers are therefore not equally distributed: Lombardy has almost half of the sample, and this makes sense considering that Lombardy is the region with the most economic activity.

Looking at the pool of taxpayers used for all analyses (the study uses all subjects to estimate the presumptive income of G1), the subjects are distributed among the different sectors and regions, as Table 2 shows.

Table 2.2: Distribution of subjects by activity, 2007.

Sector	**Lombardy**	**Lazio**	**Sicily**	**Total**
Industry	17,407	4,899	6,836	29,142
Commodities	0	0	0	0
Water	121	42	40	203
Construction	39,560	10,907	9,677	60,144
Commerce	59,154	31,790	30,726	121,670
Transportation	8,669	2,188	2,250	13,107
Hotels & restaurants	8,752	5,328	4,899	18,979
Information	5,063	2,007	563	7,633
Finance	3,463	1,601	932	5,996
Real estate	3,110	1,381	457	4,948
Professional	65,950	34,677	17,595	118,222
Holidays	5,530	2,180	1,268	8,978
Public admin.	1	3	1	5
Education	185	139	147	471
Healthcare	12,698	5,237	3,313	21,248
Arts	1,221	1,952	378	3,551
Other services	11,538	4,384	2,914	18,836
Servants	0	0	0	0
International org.	0	0	0	0
Total	242,422	108,715	81,996	433,133

Most taxpayers are concentrated in a few sectors (commerce and other services), while some other sectors can be considered marginal. The following section illustrates the two methodologies to quantify tax evasion. These methodolo-

gies will correct the reported incomes for these sectors and provide an estimation of the evaded incomes.

2.5.2 Group 1 characteristics

This study focuses on the G1 group, which includes mostly small entrepreneurs and professionals. G1 contains taxpayers where the owner's work is similar to that of a dependent worker. These units are characterized by a very low level of capital and the largest input in the production function is the individual's personal work, while the use of workers or fixed assets is seen as an obstacle to freedom of activity. The main characteristics are:

- Not being a limited company or controlled by a limited company;
- Not having non-working members;
- Adheres to the minimum tax scheme;
- Declares total revenues less than € 30,000;
- Independent workers hired less than 1.5 but higher than zero;
- Dependent workers hired virtually null, less than 0.5;
- Being older than 30 years.

There are multiple reasons for choosing this group. First, this study should focus on analyzing taxpayers that have a real possibility of tax evasion. For this reason, this database, which includes only self-employed taxpayers and sole proprietorships, is a good choice. This group of taxpayers has the possibility to deliberately hide part or all of their economic transactions, and thus efforts to counteract evasion focus on this group.

Self-employed persons and sole proprietorships are included in different groups: among all groups, this study focuses on G1. Provided the information included

in the database, it is possible to replicate the methodology for G1. The different estimations by various institutions (such as Istat, Banca d'Italia, the RA, etc.) find evidence that this group has the minimum level of tax compliance. It is common knowledge that G1 taxpayers evade a taxes great deal, but their size is so small that the total amount of tax not declared is just a minor part of the revenues that the revenue agency does not collect. The low level of compliance make this group quite interesting to study, and the relatively limited size makes it possible to implement a unique methodology to correct under-declared incomes.

Moreover, G1 is divided into three smaller groups, which are described fully in 2.A. Table 2.3 summarizes the average incomes, standard deviation, and minimum and maximum incomes for all of the taxpayers in G1.

Table 2.3: Income summary.

Year	**N**	**AV Income**	**SD**	**Min**	**Max**
2007	23,621	17,262.67	24797.73	0	1,085,999
2009	25,130	17,105.82	30547.70	0	2,120,081
2010	25,422	17,437.90	42259.59	0	4,253,140
2011	25,381	18,225.51	57575.03	0	7,652,803

Table 2.4: Group 1 distribution of subjects by activity.

Activity	**2007**		**2009**		**2010**		**2011**	
	Tot	%	Tot	%	Tot	%	Tot	%
Industry	875	3.70	1,091	4.34	1040	4.09	1025	4.04
Construction	3,205	13.57	3624	14.42	3687	14.50	3609	14.22
Commerce	5,685	24.07	6031	24	6094	23.97	5942	23.41
Other serv.	13,856	58.66	14,384	57.24	14,601	57.43	14,805	58.33
Tot.	23,621	100	25,130	100	25,422	100	25,381	100

The table shows no significant variation in taxpayers' activities and that the most represented category is other services, while commerce and construction are less represented, and industry is a marginal choice.

Unfortunately, the information in the database does not allow an estimation of shadow incomes for the other groups. As taxpayers increase in size and complexity, the methodologies to estimate shadow incomes become more complex, and require more variables. Sadly, no other Istat methodology could be implemented considering the variables included in the database. For this reason, this study provides estimations for G1 only.

2.6 Results

2.6.1 Outline

Now that the characteristics of presumptive taxation have been outlined, the possible reactions described, and the differences between the two methodologies illustrated, it is possible to quantify the magnitude of the introduction of the minimum tax.

This section aims to shed some light on the effects of introducing minimum taxation, particularly looking at the variation in the tax collected, the changes in the concentration, and the differences in taxpayers' participation in the market.

The main purpose will be, then, to provide a more precise description of the intuitions introduced earlier. Prior studies show that introducing a minimum tax causes an increase in the tax collected, but potential taxpayer reactions (the leaving and levelling effects) might alter the results of the base case.

Since the presumptive taxes are estimated using two different methodologies to correct income, there will be a brief description of the differences between them.

2.6.2 Differences in the income corrections

Since this study uses two methodologies to calculate the presumptive incomes that form the tax base for presumptive taxation, it is important to highlight the differences between them. The previous section illustrated that these methodologies have different purposes and characteristics. Thus, the variations in the estimation of presumptive incomes are a direct consequence of these different characteristics. The Istat presumptive statistical income was estimated following the Istat methodology to estimate shadow income, and the results for the observed years are shown in Table 5.

Table 2.5: Income correction: Istat methodology.

Year	Declared	Presumptive	%	Istat Official
2007	335,084,142	530,516,033	158.32%	
2009	359,081,559	706,243,593	196.68%	
2010	373,081,476	668,567,155	179.20%	178.50%
2011	387,303,089	687,194,856	177.43%	162.60%

The presumptive incomes are significantly higher than the actual incomes, and this is true for all years. Moreover, Istat provides an official estimation for G1 for 2010 and 2011. As Table 6 shows, the calculated presumptive incomes are quite close to the Istat official values, validating this work. The database includes the RA's presumptive operational income. Table 6 compares the actual incomes.

Table 2.6: Income correction: RA methodology.

Year	Declared	Presumptive	%
2007	335,084,142	474,117,969	141.49%
2009	359,081,559	507,661,835	141.38%
2010	373,081,476	484,345,553	129.82%
2011	387,303,089	487,154,145	125.78%

2.6.3 Differences in the correction

Starting with the initial case, where there are no reactions to the introduction of a minimum tax, as specified by 2.4, we expect the total amount of tax collected to increase and the disproportionality described by 2.6 to decrease.

If we consider an initial distribution of taxes and incomes declared by taxpayers as in 2.3, it is easier to observe an improvement after implementing a minimum tax. Comparing the initial state with the images showing the distributions of taxpayers whose tax base was modified in consideration of both methodologies in 2007 [4]. We expect to see an increase in the tax collected for taxpayers with a low declared tax base. Thanks to the graphical representation, it is possible to see the effect of introducing a presumptive tax.

In all the figures below, the order of taxpayers on the x axis represents the declared tax base. By not changing the order in the x axis, it is easier to appreciate the differences after changing the settings. The dots represent the declared income and taxes paid by each taxpayer. The changes in the position of the dots represent how implementing a minimum tax changes the taxes due while incomes remain the same.

If the Istat presumptive tax is implemented, several taxpayers will pay more and the differences in the scatter representation of the two images show an increase in the total amount of tax collected. We observe a shift upward of the distribution of the taxpayers, meaning citizens are paying more taxes. Moreover the distribution of taxpayers changes, suggesting that the proportion in taxes paid also differs. It is possible to observe that a shift upward in the taxes paid, especially for taxpayers with incomes below € 30,000.

[4] For graphical reasons, the next figures include only the taxpayers with a total tax base (also including non-evadible income) of less than € 100,000 and taxes lower than € 40,000

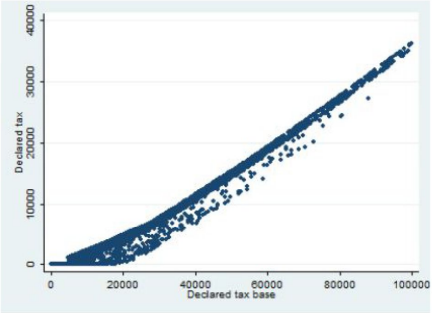


Figure 2.3: Declared taxes

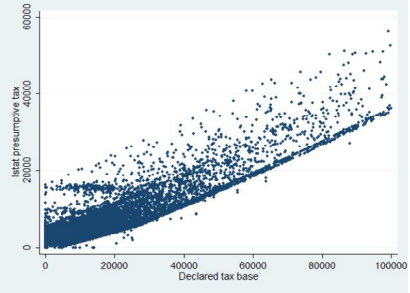


Figure 2.4: Istat presumptive taxes

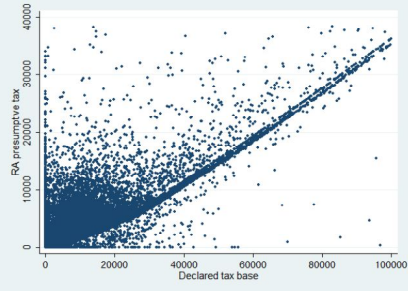


Figure 2.5: RA presumptive taxes

It is possible now to determine whether the same results are obtained using the RA's methodology. If we compare the introduction of the RA's presumptive taxes (Figure2.4) to the declared taxes (Figure2.3), we can draw the same con-

clusions as for the Istat methodology. Introducing a minimum tax will increase the total amount of tax revenue and will cause a change in the coefficient of taxation.

Specifically, it is possible to compare the effect of introducing presumptive taxes using the two different methodologies.

Both methodologies increased the tax collected compared to reported taxes. However, the two images show an upward correction in declared taxes, but also show unequal amounts in the corrections.

The previous representations gave the insight that introducing a minimum tax caused an increase in the total amount of tax collected. This insight was unambiguous from the results using both methodologies, although there were some differences between the graphs.

Table 2.7: Changes in revenues and concentration with two different methodologies

Istat	**2007**	**2009**	**2010**	**2011**
Δ revenues	61.2 %	114.5%	87.6%	84.9 %
Δ concentration	-27.9 %	-57.8%	-49.7%	-46.6 %

RA	**2007**	**2009**	**2010**	**2011**
Δ revenues	45.7 %	50.2 %	29.7 %	25.6 %
Δ concentration	-26.0 %	-29.7 %	-29.9 %	-25.4 %

The results clearly show that for both methodologies without reactions, there is a significant increase in the tax collected and a decrease in the variation in the concentration. There are some differences for each year, but the general tendency shows that all of the variations are in the expected directions.

Specifically, it appears that the two methodologies differ not only in terms of the total amount of revenues collected when implementing a minimum tax, but also some differences in the variations of the concentrations: the Istat method-

ology provides a higher change in the concentration. This happens because this methodology corrects taxpayers declaring incomes much lower that the presumptive incomes more strongly. The RA methodology does not correct these taxpayers so much, and thus the RA correction preserves more of the disproportionality that was present before the minimum tax implementation.

If the government introduces a minimum tax, and leaving and levelling effects are absent, there is a remarkable increase in the tax collected. In the simulations with the taxpayers included in the database, the total increase in the collected revenues ranges from 25% to 114%. Moreover, the increase in the tax collected occurs with both methodologies, although the Istat method brings higher increases than that of the RA method.

Fig. 2.6 represents the effect of the distribution of the fiscal burden among taxpayers. The analysis focuses on the results for both methodologies for 2007 only. As the previous table suggests, the results for all years are very similar; thus, a single representation of both methodologies using Lorenz curves (with the population on the horizontal axis ordered by declared income and taxes paid on the y axis) helps to clarify the main dynamics for all observed periods.

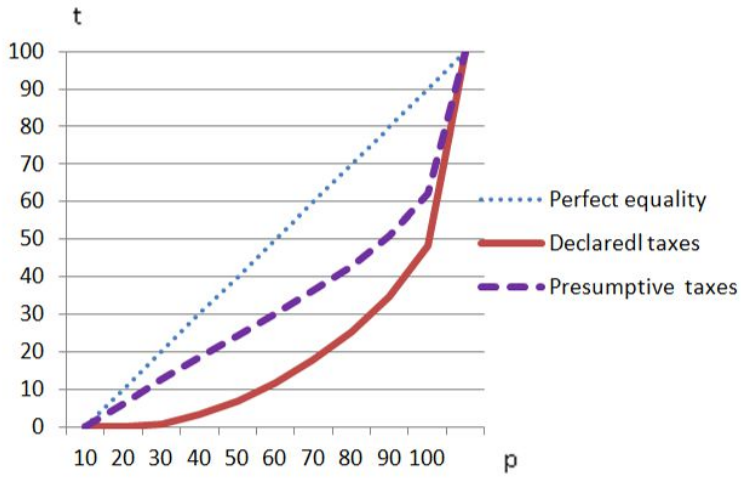


Figure 2.6: Comparison of taxes before and after the minimum tax implementation, Istat methodology

Considering that the x axis refers to the population ordered by declared taxes, it appears that most of the upward correction concerns low-income declaring taxpayers. Using Lorenz curves helps to show that introducing a minimum tax will move the Lorenz curve of the fiscal burden towards the line of perfect equality, signaling a new, lower difference in the tax collected. Thus, adopting the Istat methodology, $T_D^j < T_P^j$ taxpayers pay more taxes, and subjects with the lowest declared taxes are corrected the most.

If the same exposition is proposed for the RA methodology, we can observe similar results. The results in Table 2.7 confirm that for the RA method as well, introducing a minimum tax will increase the total amount of tax collected and reduce the concentration of the fiscal burden. Moreover, as with the Istat methodology, the RA methodology brings similar results for all observed years. Provided these characteristics, fig. 2.7 represents only 2007 since the analysis

is similar for all years and can be extended to the other periods without loss of generalizability.

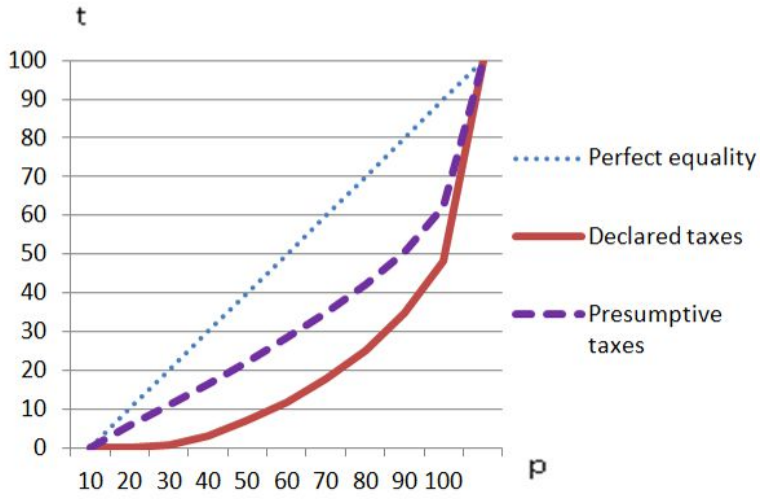


Figure 2.7: Comparison of taxes before and after the minimum tax implementation, RA methodology

If a minimum tax is used, as in (6), it appears that presumptive taxes make taxpayers with low declared incomes pay more taxes. Since this result is very similar to the Istat result, the same conclusions can be drawn. Introducing a minimum tax will affect taxpayers with low declared incomes more.

2.6.4 Effects of reactions to the minimum tax introduction

The previous section explained that introducing a minimum tax may cause some reactions from taxpayers. Recalling the earlier illustrations, there can be two different effects: a leaving effect and a levelling effect.

If we consider the outcomes by analyzing different levels of sigma and maintaining the order of taxpayers' declared tax base on the x axis, it is possible to

analyze the reactions in terms of the leaving and levelling effects.

Moving from mild to strong reactions, we can analyze the scatter distribution of all taxpayers. In these three figures, the x axis has the declared tax base to keep the same order in the distribution of the subjects, while the y axis shows the taxes declared after the subjects are aware of the presumptive taxes and can pay based on their level of reaction.

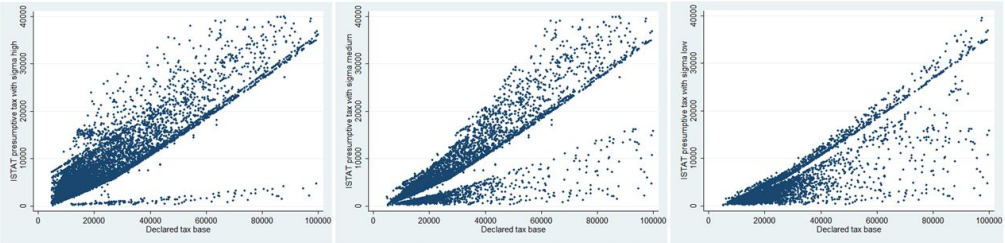


Figure 2.8: Istat presumptive taxes with high, medium, and low σ

As σ changes, the remaining subjects are almost divided into two groups: a higher and a lower one. In the higher group, we have subjects in groups 2 and 3, while in the lower group, we have subjects in group 4. It is possible to observe that as σ decreases and the reactions become stronger, the upper group of subjects decreases (as groups 2 and 3 shrink), while we find that the lower group includes more subjects (group 4 increases). Considering these effects together, the total amount of tax collected will be lower since there are fewer taxpayers in the market (due to the leaving effect) and more remaining subjects will pay less taxes (the levelling effect). At the same time and for the same reasons, the coefficient of concentration should also be affected since as σ decreases, the coefficient of concentration should increase.

Unfortunately, the figures above do not allow for a clear representation of

the subjects leaving the market. Since there are more than 20,000 observations, it is necessary to provide a representation of these subjects separately better understand them because they change with different levels of σ. In the next figures, we observe the leaving effects considering that the y axis shows the presumptive taxes of subjects who decide to leave the market. The figures then show the taxes that are not collected and the negative effects of a high level of reactions to the minimum tax implementation.

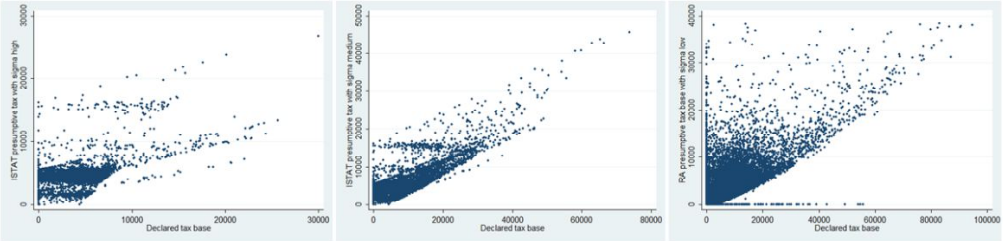


Figure 2.9: Istat methodology: Taxpayers leaving the market when $T_D^j < T_P^j$ with high, medium, and low σ

These figures show the scatter representation of the presumptive taxes of subjects leaving the market for three σ settings. As σ decreases, more taxpayers consider the presumptive taxes too expensive and decide to leave the market.

Now, it is possible to analyze the levelling effect. If we look at the subjects in Sector 4, with declared taxes below € 15,000, we obtain the following figures.

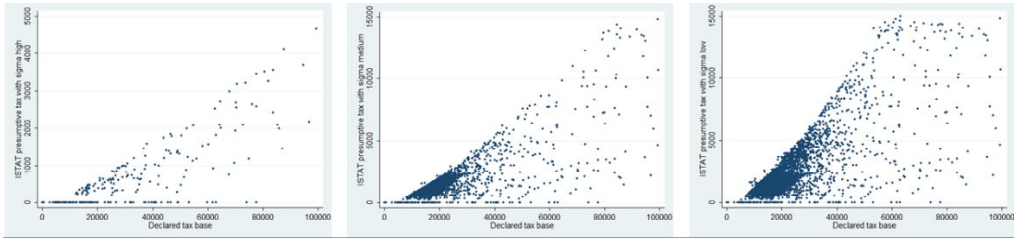


Figure 2.10: Istat methodology: taxpayers paying presumptive taxes when $T_D^j > T_P^j$ with high, medium, and low σ

We see that as σ decreases, the number of subjects included in group 4 increases. This means that more taxpayers are paying a lower level of taxes since their declared taxes are lower than the presumptive ones, causing a decrease in the total amount of tax collected.

Considering all figures together, it appears that an increase in taxpayers' reactions decreases the total amount of tax collected, since more taxpayers leave the market or declare presumptive taxes. Moreover, there are also changes in the concentration of taxes paid. Due to the leaving and levelling effects, fewer taxpayers are in the market and the remaining participants will pay similar taxes as σ decreases, causing a change that will be measured by the coefficient of concentration.

The next figures show how reactions will affect the final outcome using the same levels of reactions but with the RA methodology. As before, the exposition of the results will focus first on taxpayers' overall responses, while the differences among the groups will be investigated secondly.

The following figures show the total effect with different levels of σ, starting with the case of low reactions (σ high) and moving to strong reactions (σ low).

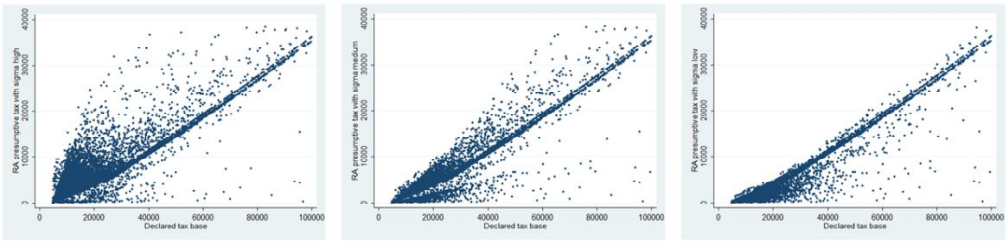


Figure 2.11: RA methodology: presumptive taxes with high, medium, and low σ

It is possible to observe that now the subjects are less dispersed and, as σ decreases, there is a consequent increase in the number of the subjects below the line, where most of the taxpayers lie.

The changes are caused by the reactions becoming stronger. As σ decreases, more taxpayers leave the market or pay presumptive taxes. As the images show, the total amount of tax collected decreases while the change in the coefficient of concentration moves in the opposite direction.

To clarify the leaving and levelling effects, it is better to analyze taxpayers in the same way as for the Istat methodology.

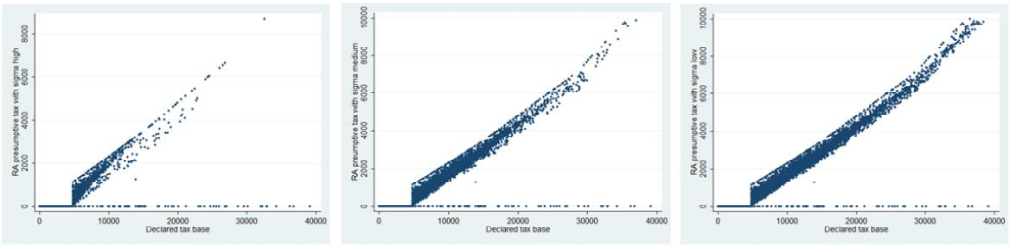


Figure 2.12: RA methodology: taxpayers leaving the market when $T_D^j > T_P^j$ with high, medium, and low σ

Observing how taxpayers with $T_D^j < T_P^j$ behave, when reactions are stronger, more taxpayers consider that paying the presumptive tax is too expensive. Group 1 increases and more taxpayers leave the market.

Finally, it is possible to observe what happens in group 4, where subjects pay presumptive rather than the actual taxes when $T_D^j < T_P^j$.

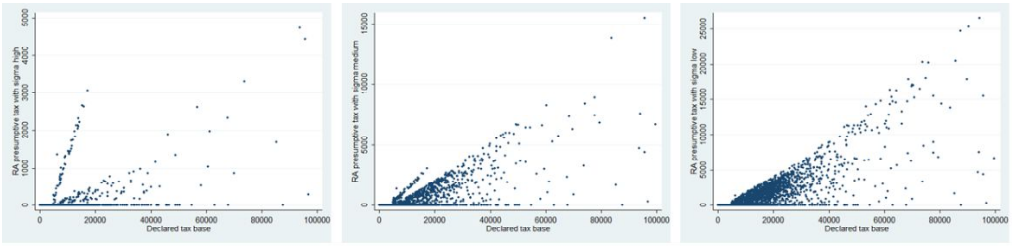


Figure 2.13: RA methodology: taxpayers paying presumptive taxes when $T_D^j > T_P^j$ with high, medium, and low σ

Here, we can see an increase in the number of the subjects as σ increases. The levelling effect makes more subjects move from group 3 to group 4, paying presumptive taxes rather than declared taxes, causing a fall in the total tax collected.

It is possible to draw the same conclusions, as the Istat methodology are duplicated for the RA methodology, although the figures shows different scatter distributions due to the differences in the methodologies. Stronger reactions are associated with decreases in the total amount of tax collected and in the concentration of the tax paid. What is crucial is that the levelling and leaving effects are present in both methodologies, and the next section provides a more precise quantification of these effects.

To analyze the combined effects of these two reactions, it is possible to have a look at the total tax collected and at the concentration among the taxpayers.

The figures in the previous section already show the results of the effects in three different contexts: strong, intermediate, and mild taxpayer reactions. It is thus possible to observe the main dynamics, understand the roles of taxpayer reactions, and quantify the final effects.

Starting with the Istat methodology, we want to observe how the total taxes collected changes as the reaction becomes stronger, how many taxpayers continue to stay in the market, and how the concentration of taxes paid changes.

Table 2.8: Changes in revenues and concentration: Istat methodology with different levels of σ.

Istat	2007			2009		
	Low	mid	High	Low	Mid	High
Δ revenues	-53.70%	-10.80%	35.00%	-57.50%	-17.50%	43.60%
Δ concentration	-3.30%	-11.20%	-22.70%	-3.90%	-13.70%	-32.20%
Δ population	-64.20%	-45.20%	-20.50%	-69.50%	-54.00%	-29.00%
	2010			2011		
	Low	mid	High	Low	Mid	High
Δ revenues	-43.09%	-16.75%	13.32%	-44.67%	-16.28%	19.68%
Δ concentration	-8.62%	-14.44%	-25.24%	-3.20%	-12.47%	-28.56%
Δ population	-64.29%	-51.88%	-36.17%	-67.21%	-54.29%	-34.58%

The results show that as taxpayers' reactions become stronger (from σ low to σ high), we experience a severe loss in market participation. It is also possible to notice that as the reactions increase (lower σ) there is an increase in the concentration of taxes paid. By a change in the concentration, we mean that the difference between the Gini index of the taxes paid and the concentration index of taxes, in the same order as that of the taxes paid.

If we want to control the differences in the distribution of taxes, it is possible to represent the Lorenz curves with different levels of σ. In Figure 14, the x axis is the same order of declared taxes, while the y axis shows the different taxes: actual taxes without reaction and taxes paid with different reaction levels. Note that this figure excludes the subjects that decided to leave the market.

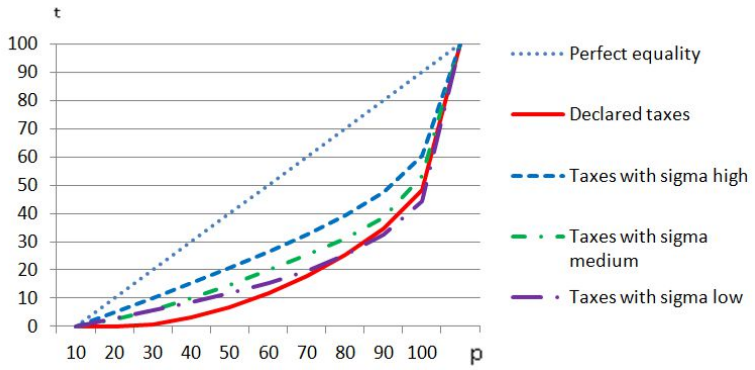


Figure 2.14: Comparison of distributions of the fiscal burden with different reaction levels, Istat methodology.

The graphical representation shows us that introducing different levels of reaction brings important changes to the distribution of the fiscal burden. If reactions are limited (σ high), we observe that the fiscal burden is more homogeneous among taxpayers: from the Lorenz curve, we move towards the line of perfect equality. However, as the reaction become stronger, we observe that the

Lorenz curves move away from the line of perfect equality due to the leaving and levelling effects.

Once we see the main dynamics for the Istat methodology, it is possible to focus on the effects of the same reactions on the introduction of a minimum tax with the RA methodology.

Table 2.9: Changes in revenues and concentration with RA methodology and different levels of σ

RA	2007			2009		
	Low	mid	High	Low	Mid	High
Δ revenues	-27.90%	-7.30%	19.90%	-22.80%	-8.90%	11.60%
Δ concentration	-4.80%	-8.10%	-18.30%	-7.30%	-11.00%	-23.50%
Δ population	-46.30%	-35.60%	-20.50%	-40.50%	-33.90%	-23.00%
	2010			2011		
	Low	mid	High	Low	Mid	High
Δ revenues	-7.90%	-0.40%	10.90%	-8.70%	-0.50%	11.10%
Δ concentration	-10.30%	-12.40%	-18.30%	-11.90%	-14.00%	-20.10%
Δ population	-32.60%	-27.90%	-20.40%	-31.80%	-26.70%	-18.60%

The representation of the correction is the same as for the Istat methodology. The table represents the effects of introducing a minimum tax with three different reaction levels and the results are proposed for all years available.

It appears that the results are very similar to those of the Istat method. As reactions become stronger (from σ high to σ low), fewer taxpayers participate in the market, less resources are collected, and the variation in the concentration increases. Although the results are different, we observe that the main dynamics are the same as for the Istat methodology.

It is now possible to use the Lorenz curves to illustrate the distribution of the fiscal burden and see how this distribution changes with different reaction levels.

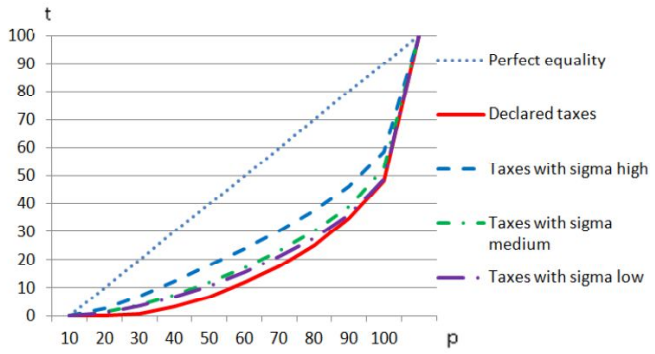


Figure 2.15: Comparison of the fiscal burden distribution with different reaction levels and the RA methodology.

Although the Lorenz curves change, meaning that the correction using the RA methodology differs from the Istat correction, the main dynamics are the same as for Istat. Introducing a presumptive tax with a low reaction level (σ high) brings a higher level of equality, and so the Lorenz curve is closer to the line of perfect equality than that for declared taxes. However, as the reaction becomes stronger, the leaving and levelling effects change the concentration coefficients, causing the Lorenz curve to move away from the line of perfect equality.

In conclusion, it is possible to observe that the correction from the two methodologies also changes when we introduce reactions. It appears that the effect of increasing reactions affects the change in revenues and the population differently, since the Istat methodology gives more extreme results than the RA methodology. Regarding the change in the concentration, this study finds evidence that the Istat methodology is characterized by a remarkable change when reactions are weak. However, as reactions strengthen, the change decreases compared to the RA methodology, meaning that a large part of the correction already occur when reactions are low.

2.6.5 A different representation

It is possible to provide a different representation of the distribution of the taxpayers. If we consider their in the vertical axe still the presumptive taxes while in the horizontal one the declared ones (instead of the tax base) the relationship between the two taxes can be analyzed in a different perspective.

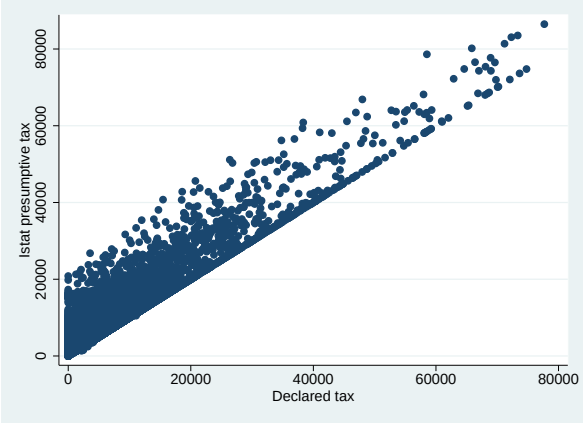


Figure 2.16: Relationship between declared taxes and Istat presumptive taxes

Starting with the Istat methodology, we observe in fig. 2.16 that the presumptive taxes bring an increase of the amount of tax collected. The same representation can be obtain using the RA presumptive tax.

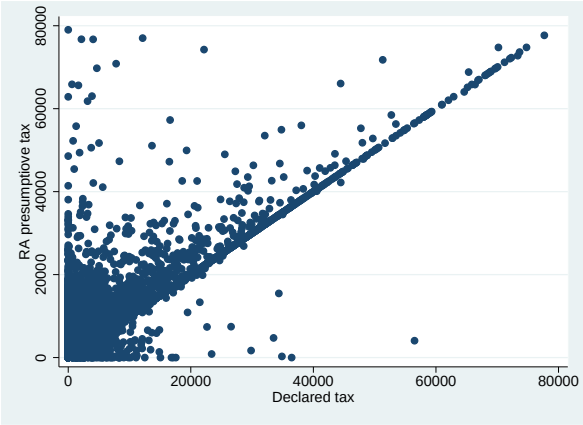
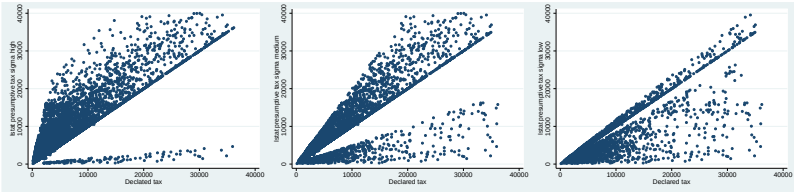


Figure 2.17: Relationship between declared taxes and RA presumptive taxes

Also fig. 2.17 shows an increase in the total amount of tax collected, but the correction seems to be more disperse than the previous one. Also this different representation catch the differences between the two methodologies. As reported before the Istat methodology corrects taxpayers declaring incomes much lower that the presumptive incomes more strongly. Instead the RA methodology does not correct these taxpayers so much, and thus the RA correction preserves more of the disproportionality that was present before the minimum tax implementation as 2.7 measured..

Once reactions are included, it is possible to observe how the relationship between the declared tax and the presumptive tax changes. As before the istat methodology is analyzed first.

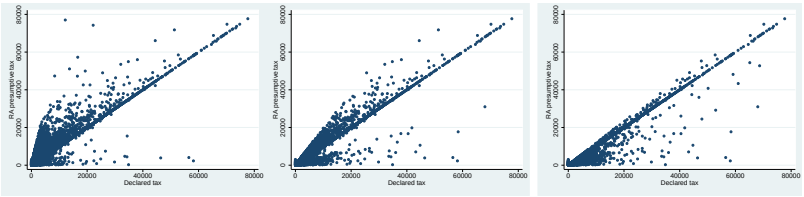


(a) (b) (c)

Figure 2.18: Relationship between declared taxes and Istat presumptive taxes with *sigma* high, medium and low

If the reactions of taxpayers change as in fig. 2.18, from a low level of reaction (a) to a high level of reaction (c), more taxpayers declare lower taxes. Since more taxpayers with $T_D^j > T_P^j$ are choosing to pay the presumptive taxes and so they are passing in the lower part of the bisector. As before above the bisector we have the taxpayers included in sectors 2 and 3 while taxpayers in sector 4 are in the lower part. If the level of reaction moves from a mild level to a strong one taxpayers in sector 2 leave the market, while taxpayers in sector 3 move to sector 4.

The same dynamics arise if we consider the RA methodology. As for Istat, we select three different levels of reactions and observe how taxpayers behave.



(a) (b) (c)

Figure 2.19: Relationship between declared taxes and RA presumptive taxes with *sigma* high, medium and low

Also in this representation, moving from a low level of reaction (a) to a high level of reaction (c) the number of taxpayers above the bisector are decreasing.

The differences with the Istat simulation are due to the peculiarities of these methodologies The same dynamics that are described for the Istat methodology apply for the RA one, in a different way though.

2.6.6 Trade-off between revenues and concentration

It is now possible to focus on the effects of introducing a minimum tax on revenues and concentration. After the analysis, it is possible to focus on Yitzhaki (2007) and his predictions.

Adopting both methodologies without reactions from taxpayers, we find that introducing a minimum tax increases the total revenues in all years analyzed. Although there are differences between the two methodologies and the yearly results, introducing a minimum tax increases the total amount of tax collected significantly compared to the actual revenues. This increase ranges from 25.6% to 114.5%, meaning that introducing a minimum tax has an unequivocal positive effect on tax revenues.

Another effect is the change in the concentration of taxes. Considering the various years and methodologies, there is an unambiguous decrease in the variation in the concentration. As for the increase in revenues, the results are quite varied and range from -25.4% to -57.8%; thus, the effect is also clear for the concentration of taxes.

If taxpayers can react to the introduction of a minimum tax, it is possible to observe that the total amount of revenue collected is higher than from the declared tax, assuming a low level of reaction (σ high). At the same time, it is possible to observe that a limited number of taxpayers leave the market, but the variation in the concentration is important. If reactions are strong (σ low), the total amount of tax collected is inferior to the total amount of tax declared, a high number of taxpayers leave the market, and the variation in the

concentration is quite contained.

Observing the results, it emerges that with moderate reactions (σ medium), the total amount of tax collected is almost the same as the declared taxes, but there are two differences: fewer people participate in the market and there is a significant difference in the distribution of taxes.

2.7 Conclusions

The analysis suggests that introducing a minimum tax would be not convenient when taxpayers begin to have significant reactions. When reactions become considerable, the total amount of collected taxes will be less than the total amount of declared taxes, but with a lower level of participation in the market and a continuing disproportional effect, although it is limited. If taxpayers' reactions are limited and there is a trade-off between higher revenues, then limited variation in participation and a significant variation in concentration should be considered.

In a context of limited reactions, implementing a minimum tax can be justified by lower compliance costs for taxpayers and lower costs for the public administration. Among the reasons Yitzhaki (2007) identifies, it emerges that minimum taxes are mostly used in underdeveloped countries, since their implementation is quite easy and the only way for states with limited resources to invest in a modern tax system. Moreover, minimum taxes can also be a viable solution for some developed countries. If compliance costs are high and it is too costly for a revenue agency to run book audits, implementing a minimum tax can be a good solution for financing the state in an economical way.

If limited reactions characterize Italy, implementing a minimum tax will be a viable policy. Since some reports [5] consider the Italian tax system as one of

[5] Among, these "Doing business" by Price Waterhouse Coopers.

the most costly one in terms of compliance, introducing a minimum tax would dramatically lower these costs. Moreover, it would be much easier for the RA to administer the application of a minimum tax compared to book audits. Then, a minimum tax would help the government recover uncollected revenues due to the high level of tax evasion that characterizes Italy, and the cost as a result of the change in concentration would be tolerable.

Considering that each single audit costs to RA €1,925 (see Mazzolini (2017)), the implementation of a minimum tax in a context of limited reactions requires just an investment of €12,849,375 in the context of the Istat methodology and of €13,240,150 in the context of RA to audit all the taxpayers in group 4. Since the implementation of a minimum tax will cause a predicted increase of the collected taxes equal to €32,072,869 with the Istat methodology and to €18,212,188 with the RA one. Unfortunately, there is no information available on the resources invested by RA to audit G1 to make a comparison.

However, it is not guaranteed that taxpayers' reactions would be limited. If Italian taxpayers have a small reaction, introducing a minimum tax is a viable policy to solve the problems reported above (high levels of evasion, high compliance costs, costly audits given the RA's resources), but high levels of reaction would worsen the initial situation. As shown above, if the government introduces a minimum tax in a context of moderate reactions, the outcome is not desirable since it would bring lower revenues and changes in concentration. Only in the presence of limited reactions can the considerable decreases in compliance and administrative costs justify the trade-off between revenues and concentration changes that a minimum tax brings.

Future developments of this work might study different solutions to the problem of the correct perception of presumptive taxes as minimum taxes. This study measures reactions by σ and focuses on the effects of different levels of

reactions. It will be interesting to analyze changes in the policy aiming to influence reactions. A promising example is the possibility of considering the implementation of credible audits on the taxpayers with $T_D^j > T_P^j$ to enable them to comply to the minimum tax policy. Moreover, it would be interesting to endogenize the control costs to investigate the possibility of limiting the levelling effect.

Further, it will be possible to investigate various solutions to limit the leaving effect. One solution could be to introduce a progressive minimum tax to avoid taxpayers leaving the market. Another possibility would be to investigate other proxies of their tax base or the possibility of finding other assets to tax. Finally, it would also be possible to analyze a taxpayer subsidy to help them to stay in the market. In this case, it is only necessary to provide with adequate resources to reach the threshold level at which they are indifferent between staying in the market or leaving it.

Appendix 2.A Groups classification

The new methodology developed by the Italian Statistical Institution (Istat) in 2014 [6] is characterized by an increase in the number of economic subjects analyzed and in the precision and conceptual distinction between the phase of finding the underreporting subjects and revising the declared income.

The taxpayers included in the analysis are those with fewer than 100 workers[7] and outside the agricultural or related sectors[8]. The observed subjects are then analyzed separately in homogeneous groups. More specifically, using the economic characteristics included in the Frame-Sbs database[9], Istat divides the

[6] A.Puggioni et al. "L'Economia Non Osservata nei Conti Nazionali" (2015)

[7] This is the result of the data analyzed by Istat. Only information on firms with fewer than 100 workers are collected by this institution for the Frame-Sbs database.

[8] It is worth recalling that incomes in these sectors are calculated on a forfeit base.

[9] Frame-Sbs is the database that Istat used in the work. In greater detail, it includes

pool into the following macro-classes:

- Group 1: minimal dimension taxpayers;
- Group 2: micro units;
- Group 3: organized units;
- Group 4: national groups units; and,
- Group 5: taxpayers unable to be analyzed.

Group 1 includes mostly small entrepreneurs and professionals. This group comprises taxpayers for which the work of the owner is similar to that of a dependent worker. These taxpayers are characterized by very low capital. The largest input into the production function is the personal work of the individual, whereas using workers or fixed assets is viewed as an obstacle to freedom of activity. The main characteristics are as follows:

- Not being a limited company or a being controlled by a limited company;
- Not being a non-working member of any firm;
- Adhering to the minimum tax scheme;
- Declaring total revenues less than €30,000;
- Having less than 1.5 independent workers hired but more than zero; and,
- Having virtually no dependent workers, at less than 0.5.

the most important economic information on all working entrepreneurs/firms (approximately 4.4 million units). The database does not include subjects engaging in agricultural or mining activities. The information is from different administrative sources and is used in a hierarchical manner. This new database was developed to increase the precision of the analysis by including more sources of information, treating the sources in a different manner, and increasing the number of populations/subpopulations (by dimension, activity sector, experience, and others). It is worth noting that the core of the database is still represented by the Emens. Emens are monthly declarations that employers must submit to report to INPS (the public Italian insurance office), including full information on wages, working hours, and other items).

Moreover, Group 1 is divided into three smaller groups. The first, G1A, includes subjects in the condition of economic marginality and who have income other than that recorded here. Possible examples include subjects earning a pension and who nevertheless continue to work to increase their annual income. Clearly, for these subjects, work generates a secondary income and, for this reason, they should be analyzed separately. This sub-group includes subjects older than 70 years (at this age, people are surely entitled to earn a social pension from the Italian government) who are assumed to have insignificant annuities.

The second sub-group, G1B, includes subjects considered to have low income-earning ability. They work in fields requiring no particular skills or training and that do not hire external workers. The entrepreneurs included in this sub-group are aged between 30 and 70 years.

The third group, G1C, includes subjects working in activities characterized by a high level of skills. This group includes subjects aged between 40 and 70 years and who work in the following sectors:

- business intermediation;
- software production;
- informatics consultancy;
- information technology services;
- auxiliary activity to the financial sector;
- insurance;
- real estate;
- scientific and technical professions;

- business support activities; and,
- medical.

These subjects are differentiated from those included in G1B by the fact that they use higher skills; however, their income is still earned mostly from their work and their fixed assets are considered negligible.

Group 2 includes micro-entities characterized by proper organization and structure, although small. The entrepreneur's work is not the only main production factor, and fixed assets are also used to produce the final product. The characteristics of this group are similar to those of the previous group, apart from the number of workers. Group 2 includes firms with 10 dependent workers if their sector of activity is in industry, or six if they operate in services.

Group 3, organized units, includes firms with a higher number of employed workers than the micro group, but that are still smaller than 100.

Group 4, national groups units, is similar to Group 3 and includes entities with more than 10 dependent workers if their sector of activity is industry or six if they operate in services. However, this group also encompasses firms linked only to Italian firms. The main criteria used by Istat are as follows:

- No direct links with a foreign firm;
- No direct links with firms with more than 100 workers; and,
- No direct links with national groups owning a foreign firm.

Group 5 is more varied and includes several firms that must be removed from the analysis for various reasons. As with Group 1, Group 5 is divided into the following sub-groups. Group 5A includes firms assumed to be unaffected by underreporting. Possible examples are firms owned by the state or public administrations. Firms in particular sectors are also considered not affected

by underreporting. Firms operating in extractions, mines, and sewers, among others, given their characteristics, are normally considered to be compliant by definition.

Group 5B comprises firms without data that can be trusted: firms with more than 100 workers and firms with direct links to a foreign firm or national group owning a foreign firm.

Group 5C includes firms in a particular period of their life (for example, initial stages, transformation, bankruptcy). The main object of this enquiry is to set up a method to compare firms in the context of perfect competition. Including firms in such specific time periods means including entities that cannot be compared with others. For this reason, excluding them from this analysis is desirable.

The last group, Group 5D, includes firms that have particular characteristics, such as cooperatives not pursuing the highest possible profit and firms purchasing their own goods. These firms cannot be compared for the same reasons as for those in Group 5C. Because the aim of the new methodology is to compare homogeneous firms to detect underreporting ones, including these firms alters the final outcome.

Provided this classification, it is clear why taxpayers included in groups G1A and G5 are not used by Istat to estimate the hidden economy. The G1A group represents taxpayers that can deliberately underreport, but their marginality makes their inclusion not useful. All sub-groups included in G5 are taxpayers that, for several reasons, are difficult to compare with other subjects. For this reason, they should be excluded from the analysis.

Appendix 2.B The two methodologies

2.B.1 General description

Different economic methodologies have been developed over time. This appendix illustrates some of the methodologies developed by two public Italian institutions.

The methodology adopted by Istat to estimate the recent change in GDP. This methodology has been declined considering the particular characteristics of different subjects.

The estimation of the gross product calculated on all of the declarations cannot be trusted because taxpayers' incentives to underreport income are important. Because taxes are levied on declared income, a person can deliberately report less revenues or higher costs to lower the taxes due. For this reason, information on real income earned by taxpayers is difficult to obtain and adopting specific techniques is required. Istat has developed a methodology that aims to calculate a threshold level— presumptive statistical income—for comparison with declared income and for estimating the underreported amount.

At the same time, the methodology adopted by RA to develop the most precise measure of presumptive income has changed. At the beginning of the 1970s, there was a radical transformation of Italian taxation. Following the introduction of the value-added tax (VAT), a radically new approach to dealing with taxpayers was developed because, for the first time, entrepreneurs and professionals were treated at the same level as large firms. In 1982, for the first time, it was possible to fine a subject on bases beyond analytic, in other words on the validity of accounting books, but also on the bases of other elements. In 1993, a new methodology based on the evaluation of the probability of a taxpayer being not compliant was introduced. This assumption was based on a comparison between presumptive and reported income. The instrument used to

estimate presumptive income was called studi di settore (sds), which was used for the first time in 1998 [10].

The construction of sds has evolved over the years. After collecting various information on internal variables (for example, costs, structure, productive process) and external variables (for example, degree of market competition, prices) of a selected sample of taxpayers, a cluster analysis of this information estimated the presumptive income of each taxpayer by adopting multiple regression techniques. It is important to stress that this procedure is not based on a sample analysis; instead, each step includes constant research on the most sensitive variables that influence the final result.

Although the final aims of Istat and RA are similar, they use different methodologies and databases to estimate evaded income. The next section provides a detailed illustration of the different methodologies.

2.B.2 Presumptive statistical income estimation for Group G1B

Starting with G1B, the first task is to identify the subjects who underreport their income. Because this group is characterized by significant use of personal work and very few physical assets, the identification approach is largely based on the assumption that an independent worker cannot declare less income than a dependent one.

To begin, Istat stratifies its pool using the CART algorithm[11] and four stratification variables: economic sector, turnover, region, and legal form. The algorithm constructs trees and nodes to divide the population. Each node includes a pool of subjects with a specific combination of values in the stratification

[10] For further clarifications, it is possible to refers to Dal Prato (2016).

[11] The CART algorithm is characterized by the binary generation of regression trees. Some observations in a dataset may be pruned to retain the most significant ones. This process is done by considering the information gain produced by each observation.

variables.

After this process is completed, the threshold income is evaluated by considering the wage of a dependent worker in each node. This methodology is similar to that which gained scientific acceptance from Franz (1985). In this context, it is possible to see that the core of the methodology is the comparison between the declared income of a taxpayer in G1B and the income of a hired taxpayer in the same occupation.

This process is undertaken in a normalized and non-normalized manner considering the differences among the full-time workers hired (managers, clerks, and workers). The aim of this procedure is to obtain two different estimations of the presumptive statistical income:

- Maximum value between an employee's average income for the normalized cost between two distinct working profiles (managers and clerks on one side and workers on the other side); and
- Maximum value between an employee's average income for the non-normalized cost between two distinct working profiles (managers and clerks on one side and workers on the other side).

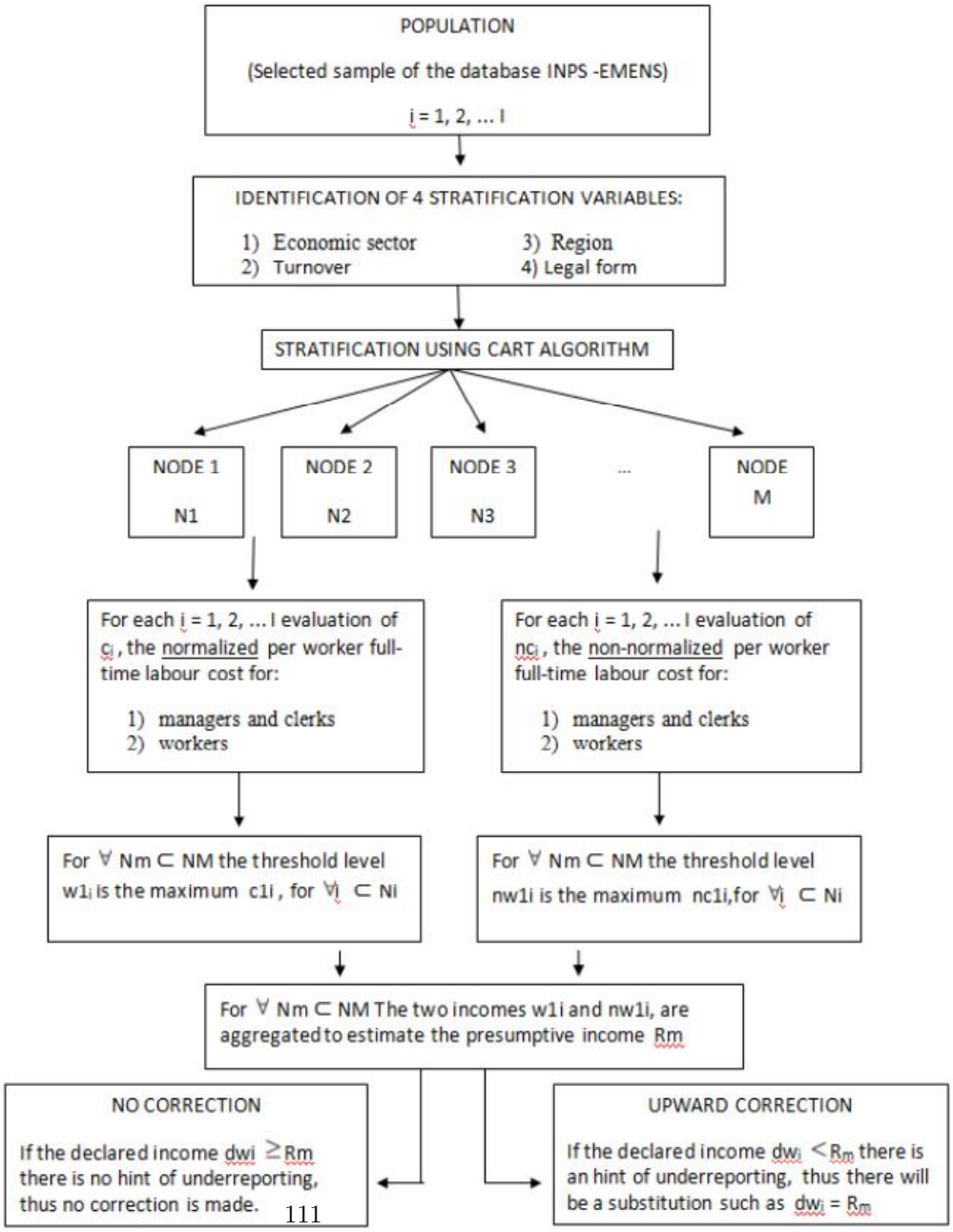
As previously stated, the starting assumption is that the income of an entrepreneur that is smaller than the correspondent one of a dependent worker is a hint that they are underreporting their income. In detail, the presumptive statistical income is estimated by considering the highest income earned by a dependent worker in each node. This comparison is made considering normalized and non-normalized incomes.

It is worth noting that the comparison of incomes is made using income before taxes. In the Italian context, this quantity is income before the IRAP tax[12], because this is the closest reported value that is representative of the cost

[12]IRAP is a tax levied on economic activities and redistributed in the region of the en-

of a dependent worker.

trepreneur/firm. This tax was established in 1997 and is the only tax directly proportional to turnover and not to annual profit.


POPULATION
(Selected sample of the database INPS -EMENS)
i = 1, 2, ... I
IDENTIFICATION OF 4 STRATIFICATION VARIABLES:
1) Economic sector
2) Turnover
3) Region
4) Legal form
STRATIFICATION USING CART ALGORITHM
NODE 1
N1
NODE 2
N2
NODE 3
N3
...
NODE
M
For each i = 1, 2, ... I evaluation of ci, the normalized per worker full-time labour cost for:
1) managers and clerks
2) workers
For each i = 1, 2, ... I evaluation of nci, the non-normalized per worker full-time labour cost for:
1) managers and clerks
2) workers
For ∀ Nm ⊂ NM the threshold level w1i is the maximum c1i, for ∀i ⊂ Ni
For ∀ Nm ⊂ NM the threshold level nw1i is the maximum nc1i,for ∀i ⊂ Ni
For ∀ Nm ⊂ NM The two incomes w1i and nw1i, are aggregated to estimate the presumptive income Rm
NO CORRECTION
If the declared income dwi ≥ Rm there is no hint of underreporting, thus no correction is made.
UPWARD CORRECTION
If the declared income dwi < Rm there is an hint of underreporting, thus there will be a substitution such as dwi = Rm

2.B.3 Presumptive statistical income estimation for Group G1C

The approach applied to analyze entities included in G1C is quite different. Although the same stratification technique is applied, the higher level of labor skills that characterize this group make it difficult to perform an identification strategy such as that for G1B. The identification strategy is a mark-up one: the aim is to build a profit function that—considering different factors—can evaluate the threshold income for each node.

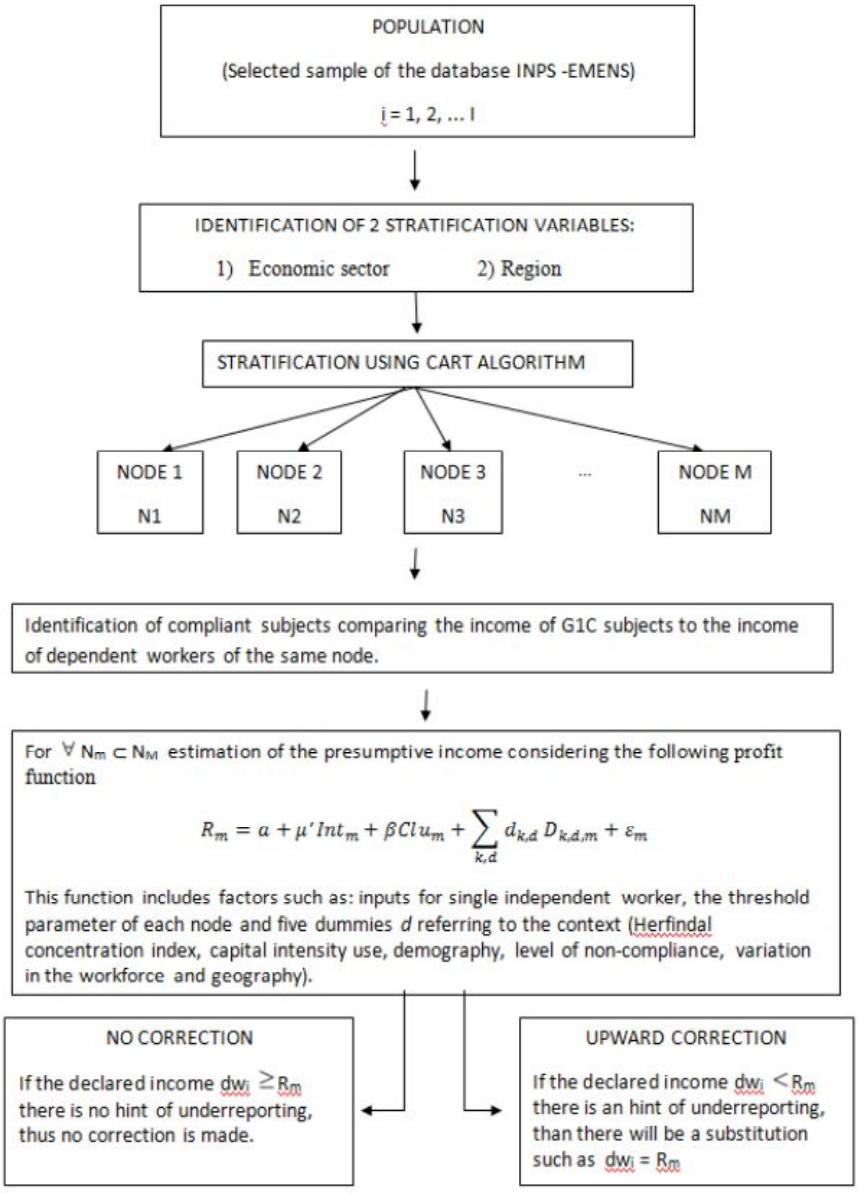
Istat stratifies its pool using only two parameters (i = sector of activity and region). The model is estimated for subjects that can be considered to be unaffected by underreporting because they have income (again, income before the IRAP tax) at least equal to the labor cost of the node.

For every node of the stratified pool "m," it is possible to represent the relationship between the entrepreneur's profit R_m*, intermediate costs Int_m, and cluster Clu_m. At the same time, it is possible to define the following multivariate linear model with dimensional effects (classes of employees), sectors (groups of economic activities), and a regional dummy $D_{k,m,i}$. The function is as follows:

$$R_m{}^* = \alpha + \mu' Int_m + \beta Clu_m + \sum_{Z_{xy} \in \vec{k,d}} d_{k,d}\, D_{k,d,m} + \epsilon_m$$

The previous function states that revenue R_m* is a function of a constant α, inputs for a single independent worker, the threshold parameter of each node, and dummies that refer to the context and error term ϵ_m. In the function, the term Int_m refers to the inputs of intermediate consumption per employee, which represents the coefficient of the mark-up. In detail, it represents the intermediate cost per independent worker. The term Clu_m represents a value peculiar to each node of the strata. This term is the cost of a dependent worker.

For this reason, it is assumed that an entrepreneur should earn at least as much as a dependent worker engaged in the same duties. Finally, there are six dummies, each distributed on a scale of five. The dummies d are the Herfindal concentration index, capital intensity use, demography, level of non-compliance, variation in the workforce, and geography (Italy has been divided into four macro areas: North West, North East, Center, and South). The dummies have a very important role in this relationship because they limit the hypothesis of strong linearity, control a large part of the heterogeneity attributable to different behaviors, and provide a more precise and accurate estimation given the values linked to the context of each entrepreneur.


POPULATION
(Selected sample of the database INPS -EMENS)
i = 1, 2, ... I
IDENTIFICATION OF 2 STRATIFICATION VARIABLES:
1) Economic sector
2) Region
STRATIFICATION USING CART ALGORITHM
NODE 1
N1
NODE 2
N2
NODE 3
N3
...
NODE M
NM
Identification of compliant subjects comparing the income of G1C subjects to the income of dependent workers of the same node.
For ∀ Nm ⊂ NM estimation of the presumptive income considering the following profit function
$R_m = a + \mu' Int_m + \beta Clu_m + \sum_{k,d} d_{k,d} D_{k,d,m} + \varepsilon_m$
This function includes factors such as: inputs for single independent worker, the threshold parameter of each node and five dummies d referring to the context (Herfindal concentration index, capital intensity use, demography, level of non-compliance, variation in the workforce and geography).
NO CORRECTION
If the declared income dwi ≥ Rm there is no hint of underreporting, thus no correction is made.
UPWARD CORRECTION
If the declared income dwi < Rm there is an hint of underreporting, than there will be a substitution such as dwi = Rm

As previously noted, once the underreporting subjects are identified, it is possible to correct the income. The correction procedure aims to substitute the income of non-compliant entrepreneurs with the presumptive income that has been evaluated for each node.

2.B.4 Italian Revenue Agency approach: Presumptive operative income

Starting from 1998, the RA has adopted a methodology to estimate the presumptive income of firms, entrepreneurs, and professionals that is referred to here as operational (to distinguish it from the Istat or statistical methodologies). Since then, different adaptations have occurred from the initial setting; however, the general layout of this instrument has not changed.

The instrument chosen to estimate presumptive income in Italy is the sds. This instrument aims to estimate the presumptive operative income of firms and professionals earning less than €7,500,000 through a step-by-step procedure that is summarized as follows:

- Identification of activities to study and relative variables by sending specific questionnaires to taxpayers;
- Analysis of the replies and removal of those considered abnormal;
- Identification of homogeneous clusters;
- Estimation of an income function for each cluster using the variables collected;
- Territorial analysis of taxpayers; and,
- Attribution of the appropriate cluster to all taxpayers.

The variables used to construct a cluster are structural (for example, number of employees, size of offices or warehouses), accounting (for example, costs of inputs), and territorial (for example, degree of competition). A detailed description of *sds* is offered by Fiorio et al. (2013), and its adaptation is provided here. RA's main aim is to create C homogeneous differentiated clusters, where each cluster c = {1, 2, ..., C} is constructed by observing the variables of the reliable taxpayers R_c, where $R_c \subseteq I_c$, and I_c is a sub-group of the total population I included in any cluster c, such that $\bigcup I_c = I$. RA estimates c relationships of revenues y' for each reliable firm in year t as follows:

$$y_{c,r,t} = \beta'_{c,t}\ x_{c,r,t} + \epsilon_{c,r,t}$$

In the equation $r \in 1, \ldots, R_c$, $x_{c,r,t}$ is the $J \times R_c$ matrix of inputs at time t, $y_(c, r, t)$ is the value of the sales reported by a reliable firm in each cluster at time t, and the last term is an error term. The aim is to estimate $\beta'_{c,t}$, the $J \times 1$ vector of unknowns for each cluster c using multiple regression techniques to bring it to $\hat{\beta}_{c,t}$. This is done because RA aims to estimate for year t the $J \times R_c$ vector of the productivity parameter coefficient: because this is done in year t+3, we have $b_(c, t + 3) = \hat{\beta}_{c,t}$.

After this vector is estimated, the presumptive sales of each firm i of the entire population of firms I still active in year $t + 3$ are evaluated as $\bar{\bar{y}}_{c,i,t} = b'_(c, t)\ x_{c,i,t}$, which can differ from the reported sales $y_(c, i, t)$. It should also be mentioned that the true inputs ($\tilde{x}_{c,i,t}$) and sales ($\tilde{y}_{c,i,t}$) can differ from the reported inputs ($x_{c,i,t}$) and sales ($y_{c,i,t}$).

Once the presumptive sales $\bar{\bar{y}}_{c,i,t}$ have been estimated for each unit, they are then compared with the declared sales $y_{c,i,t}$. If the presumptive sales are lower than the declared sales ($\bar{\bar{y}}_{c,i,t} < y_{c,i,t}$), the taxpayer is labelled congruous, to signal the positive return of this control. However, if presumptive sales are lower than declared sales ($\bar{\bar{y}}_{c,i,t} > y_{c,i,t}$), the unit is labelled incongruous and

the difference between the two sales levels (considering that costs are the same) is the hint of evasion. Being labelled incongruous has various disadvantages, such as, for example, a higher probability of being audited and facing greater difficulty receiving tax refunds.

2.B.5 Istat methodology: Results for G1B

The next step is to mimic Istat's procedure to evaluate presumptive income. Because the database is quite different from the Istat database, some adaptations are required.

The first difference concerns the stratification algorithm that has been used. In this work, the sample has been stratified using the Chi-square Automatic Interaction Detector (CHAID) stratification algorithm[13] rather than the CART algorithm. Moreover, there are fewer stratification variables than with Istat. Without information from the legal form, the stratification is done using the available information: economic sector, turnover, and region. It is worth noting that all 434,011 taxpayers have been included in the stratification and that, in the end, a total of 60 leaves were determined (see Appendix 2.B.2 for a more detailed analysis of the database).

A second difference concerns the identification of the presumptive statistical income in each stratum. The Istat methodology identifies presumptive statistical income as the highest income earned by a dependent worker in a stratum. However, Istat data are first analyzed by comparing various sources and pruning the inconsistent ones. In this manner, it is possible to investigate whether

[13]In contrast to Istat, the algorithm used in this analysis is CHAID). Unfortunately, it was impossible to run the CART algorithm with the software used (Stata 13). Among the other possible algorithms, CHAID was selected, which is similar to CART in that it assists in finding relationships among observations and analyzes the data all-in-one rather than splitting them into sub-groups and performing multiple tests. However, if CART generates a binary regression tree, CHAID builds regression trees that are not limited by being binary and that are characterized by checking whether the relationships among the variables are statistically significant.

extreme values are reasonable or are simply mistakes, allowing for corrections. Data included in this database are considered crude because such activity is not performed (the information that is reported in the fiscal forms); thus, inconsistent values are not pruned because Istat does so. To overcome this problem, instead of choosing the highest salary of a dependent worker as the presumptive income in each node, the 9th decile salary is selected. This choice is a compromise between the highest and the average such that it is still as close as possible to Istat's methodology.

Considering the previous flow chart for G1B, the stratification has been done on 434,011 taxpayers over three stratification variables (economic sector, turnover, and region). A total of 60 leaves were determined, no normalization was done, and the 9th decile of the income of a dependent worker was identified as the presumptive income.

Moreover, the database does not provide any information on the qualifications of the workforce: it is not possible to identify who is a manager or a clerk, making it impossible to estimate the normalized and non-normalized incomes of dependent workers. Identifying the presumptive income Rm is simply performed on the declared cost of the dependent workers. After Rm is identified, it is possible to compare the declared incomes to the presumptive income in each node and substitute the declared income with the presumptive one if it is smaller.

Table 2.10 shows the average incomes declared by G1B subjects in each cluster, the average of the income of the dependent workers in that cluster, and finally the average presumptive income in each cluster in 2007. It is important to state that all incomes refer to all professional incomes that are in any case referable to each unit. Possible revenues from owning land or earning rent for buildings are not included if they are not related to the profession.

Table 2.10: Calculation of average presumptive income and real income, year 2007 – Group G1B

Cluster	n.	Declared inc.	Presumpt. inc.	Av. correction	Corrected inc.
1	1,745.00	10,588.29	22,134.00	13,949.79	24,538.08
2	5,769.00	13,421.74	25,211.80	12,319.96	25,741.70
3	4,130.00	12,187.64	21,240.57	9,867.66	22,055.30
4	2,362.00	11,666.56	13,118.00	3,120.68	14,787.24
5	2,872.00	14,897.83	13,123.00	1,690.70	16,588.52
6	2,890.00	17,096.92	13,444.00	1,355.15	18,452.07
7	911	18,277.29	13,918.00	1,587.85	19,865.14
Total	20,679.00	13,668.22	18,956.46	7,435.53	21,103.75

In Table 2.10, the presumptive income is the 9th decile of the highest income earned by the dependent workers in each cluster. It is worth remembering that presumptive statistical income is the threshold level that measures whether or not an entrepreneur is supposed to be compliant. In detail, it is the minimum level that an entrepreneur has to declare to be considered compliant. If a unit of the dataset has declared income less than the presumptive income, it is considered incompliant and a correction is required. The average of the differences between the declared income and the presumptive income for each class is reported in Table 2.10: it can be noticed that the average correction is higher than the average declared income and that a large part of this correction is concentrated in Clusters 2 and 3.

After the correction, the corrected income can be estimated. For each unit, the maximum difference between the declared income and the presumptive income is taken, and then the average is quantified for each cluster. As expected, the corrected income is higher than the average declared income and the presumptive income, although the corrected income is quite close to the presump-

tive income. Because the corrected income is the maximum between the declared income and the presumptive income, it must be higher than both. However, for the corrected income to be very close to the presumptive income and remarkably different from the declared income is a hint of low compliance among the taxpayers.

Table 2.11: Estimation of average presumptive statistical income and real income – Group G1B.

Year	n.	Decl. income	Pres. Income	Pres. correction	Corrected inc.
2007	20,679	13,668	18,956	7,436	21,104
2009	21,881	13,756	20,272	8,628	22,384
2010	21,979	14,105	19,883	8,024	22,129
2011	21,895	14,806	20,816	8,221	23,026

It is possible to notice that the number of taxpayers differs for different years. This happens because some subjects fulfil the requirements to be enlisted in G1B in certain years but not in others. For example, in one year, they might employ some workers and then be entitled to be part of G2. However, the results are quite similar for all years investigated, and the conclusions drawn for 2007 can also be extended to the other years.

2.B.6 Istat methodology: Results for G1C

Focusing on G1C, the procedure is different and follows what was previously described, with some adaptations applied.

As reported in the previous section, the model developed by Istat is a mark-up model. Instead of comparing declared income to a presumptive income estimated as for G1B, the methodology developed for G1C taxpayers focuses on estimating a profit function for each cluster using the information for taxpayers who are supposed to be compliant.

As with G1B, the analysis of G1C started by a stratification of the sample. The CHAID [14] algorithm was used and, following Istat, the stratification variables were geographical location and economic activity. This resulted in 28 clusters with 13 including G1C taxpayers. Because G1C was stratified by different variables, the stratification is different from G1B, making the intermediate results not directly comparable (comparing the first clusters in the different stratifications is meaningless). Only the final results can be useful for inferences among the two groups.

$$R_m{}^* = \alpha + \mu' Int_m + \beta Clu_m + \sum_{Z_{xy} \in \vec{k,d}} d_{k,d}\ D_{k,d,m} + \epsilon_m$$

At each node level, the profits are estimated considering the intermediate costs (int) of independent workers, a parameter peculiar to each node (Clu) referring to the labor cost, and two dummies on five levels to limit the possibility of strong linearity. Given this dataset, in this study it is possible to include the dummies referred to as the Herfindal concentration index.

Moreover, to avoid collinearity, capital intensity (Cap) is not expressed as a dummy. Unfortunately, demography, level of non-compliance, variation in the workforce, and geography cannot be included in the analysis. The first three cannot be included because of a lack of data and the fourth one because of collinearity problems given that only three regions are represented in the dataset.

Another difference is in the fact that the procedure does not consider the maximum wage of the dependent worker of each node as a threshold level, in contrast to the 9th decile of Istat and the analysis of G1B. The results are too extreme if these values are used; in particular, too few subjects are compliant, making it impossible for the analysis to continue.

[14] For the same reasons as previously noted, the CART algorithm was not used for G1B; CHAID was used in the analysis of G1C.

Considering the previous flow chart for G1C, the stratification was done on 434,011 taxpayers over two stratification variables (economic sector and region). A total of 28 leaves were determined, the profit function was estimated over intermediate costs, labor costs, and two dummies (Herfindal concentration index and capital intensity), and the 5th decile of the income of a dependent worker was identified as presumptive income.

The result of the regression for 2007 is represented in the following table:

Table 2.12: Estimation of the parameters of the mark-up model, year 2007 – Group G1C.

	Coef.	Std. Err.	t	P> |t|	[95% Conf.	Interval]
Int	0.745679	0.119575	6.24	0	0.511145	0.980212
Clu	0.433694	0.304842	1.42	0.155	-0.16422	1.031608
Cap. intens.						
Herfindal	4500.561	1978.361	2.27	0.023	620.2329	8380.89
.0000455-	19588.22	3720.462	5.26	0	12290.96	26885.48
.0000911-	10116.27	3850.007	2.63	0.009	2564.921	17667.62
.0001781-						
	0.284523	0.14491	1.96	0.05	0.000299	0.568747
_cons	1741.004	4840.506	0.36	0.719	-7753.09	11235.1

The parameters are then used to estimate the presumptive income of each cluster that includes G1C taxpayers. Once the parameters are evaluated, it is possible to estimate the presumptive income for each cluster and evaluate the difference from declared income.

Regarding G1B, it is possible to summarize the main results in a table. The next table shows the average declared income, presumptive income, average correction, and corrected income for each cluster.

Table 2.13: Calculation of average presumptive income and real income, year 2007 – Group G1C

Year	n.	Decl. income	Pres. Income	Pres. correction	Corrected inc.
3	466	16,320.48	24,179.87	10,515.94	26,836.42
4	170	17,495.17	29,086.98	13,533.23	31,028.39
6	633	13,778.61	22,472.14	9,178.94	22,957.55
9	130	15,328.17	34,011.94	19,367.06	34,695.23
10	48	12,597.54	31,241.82	18,644.28	31,241.82
11	5	11,035.20	54,578.65	43,543.45	54,578.65
14	456	28,981.34	52,471.32	28,681.09	57,662.44
18	48	12,365.63	27,436.86	15,294.16	27,659.78
21	348	14,377.59	24,807.55	10,624.85	25,002.44
23	73	14,359.82	24,890.12	11,818.82	26,178.64
24	74	13,369.46	29,782.40	16,652.02	30,021.48
25	11	10,713.36	41,457.69	30,744.32	41,457.69
27	277	24,261.14	36,564.81	18,366.47	42,627.61
Total	**2,739**	**18,124.18**	**31,074.77**	**15,214.52**	**33,338.71**

The average declared income appears higher in G1C relative to G1B. This result was expected because taxpayers included in G1C are characterized by a higher degree of specialization relative to G1B. The analysis shows that by estimating presumptive income using Istat's mark-up model, the average income for this class is much higher than that of G1B. The presumptive income is €31,074.77, whereas for G1B it was estimated at €18,956.46.

Given such a difference between declared income and presumptive income, the average correction is very high—almost equal to declared income. At the same time, the corrected income is very close to the presumptive income. The results in the table show that the difference between the presumptive income and the corrected income is less than €3,000. This difference indicates that

most of the corrected income is the result of presumptive income, whereas the influence of declared income is almost negligible.

It is also possible to compare the results for G1C for 2007 with those of the other years, as provided in the following table in summary form:

Table 2.14: Estimation of average presumptive statistical income and real income – Group G1C

Year	n.	Decl. income	Pres. Income	Pres. correction	Corrected inc.
2007	2,739	18,124	31,075	15,215	33,339
2009	3,003	18,096	69,254	52,740	70,837
2010	3,145	18,684	54,203	37,879	56,563
2011	3,190	18,585	54,037	37,586	56,171

The comparison between estimations referring to G1C appear to be not as consistent as those for G1B. This result can be explained by the difference in the number of taxpayers involved in the estimation process. Provided that the estimations in G1C are based on almost one-tenth of the sample, the overall results are less accurate and more sensitive to the extreme results. Although the difference between the corrected income and the declared income can be observed to be significantly higher in all cases, a difference in magnitude exists with this divergence.

2.B.7 Istat methodology: Overall analysis of G1 and validation

Once G1B and G1C have been analyzed, it is possible to study Group 1 as a whole. First, it should be remembered that G1A was not present in the previous analysis because it was considered a group that was not evading. Because Group G1A was considered an always-compliant group of subjects, it is also possible to summarize the number of compliant and non-compliant taxpayers:

Table 2.15: Distribution of compliant subjects among classes, year 2007.

	Compliant	**Non-compl.**	**Total**
G1A	203	0	203
G1B	6,635	14,044	20,679
G1C	284	2,455	2,739
Total G1	7,122	16,499	23,621
	30.15%	69.85%	

It is possible to compare Istat's actual predictions for 2010 and 2011. In 2010, Istat estimated that 77.7% of firms included in G1 were non-compliant, whereas in 2011 the proportion was very similar, at 77.9%. The estimation carried out on the sample brought a close result because the difference is less than 10%, even if referring to a different year. In terms of a comparison with other years, the following table reports the results, showing that they are similar in all analyzed years.

To continue, it is possible to investigate the rate of correction by expressing the ratio between declared income and the average correction reported for each sub-group.

Table 2.17: Ratio of correction Group 1, year 2007.

	Declared	**Corrected**	**%**	**ISTAT 2010**	**ISTAT 2011**
G1A	2,796,809	2,796,809	/		
G1B	282,645,197	436,404,510	154.40%	199.70%	163.90%
G1C	49,642,136	91,314,714	183.95%	241.00%	247.90%
Total G1	335,084,142	530,516,033	158.32%	178.50%	162.60%

Table 2.16: Distribution of compliant subjects among classes in other years.

	2009		
	Compliant	**Non-compl.**	**Total**
G1A	246	0	246
G1B	6,485	15,396	21,881
G1C	156	2,847	3,003
Total G1	6,887	18,243	25,130
	27.41%	**72.59%**	

	2010		
	Compliant	**Non-compl.**	**Total**
G1A	298	0	298
G1B	6,959	15,020	21,979
G1C	227	2,918	3,145
Total G1	7,484	17,938	25,422
	29.44%	**70.56%**	

	2011		
	Compliant	**Non-compl.**	**Total**
G1A	296	0	296
G1B	6,603	15,292	21,895
G1C	222	2,968	3,190
Total G1	7,121	18,260	25,381
	28.06%	**71.94%**	

Table 2.18: Ratio of correction Group 1 in other years.

	2009		
	Declared	**Corrected**	**%**
G1A	3,733,656	3,733,656	/
G1B	301,004,371	489,787,759	162.72
G1C	54,343,532	212,722,178	391.44
Total G1	359,081,559	706,243,593	196.68

	2010		
	Declared	**Corrected**	**%**
G1A	4,304,293	4,304,293	/
G1B	310,015,329	486,372,970	156.89
G1C	58,761,854	177,889,892	302.73
Total G1	373,081,476	668,567,155	179.2

	2011		
	Declared	**Corrected**	**%**
G1A	3,847,184	3,847,184	/
G1B	324,168,571	504,160,892	155.52
G1C	59,287,334	179,186,780	302.23
Total G1	387,303,089	687,194,856	177.43

In addition, for the correction ratio, the results estimated on the database referring to 2007 are very close to those of Istat [15] (178.5% in 2010 and 162.6% in 2011). The other years show similar results.

[15] Istat "Il nuovo approccio alla valutazione della Noe nei Conti Nazionali" (2015).

2.B.8 RA presumptive operational income

The methodology developed by RA is different from that of Istat because of its particular mission. The main goal of RA is to find a precise estimation of each taxpayer's real income. The approach is completely different because, here, the focus is each taxpayer's income. Although Istat's classification is still used to assist when comparing the different estimations, the methodologies are completely different, as previously illustrated.

The database for this study already labeled each taxpayer as congruous or incongruous, and has the presumptive income for each subject. Without the need to mimic the estimation procedure of the institution, it is possible to analyze the pool of subjects while maintaining the groups previously identified by Istat:

Table 2.19: Distribution of compliant subjects among classes, year 2007.

	Congruous	**Incongruous**	**Total**
G1A	125	78	203
G1B	11,738	8,941	20,679
G1C	1,721	1,018	2,739
Total G1	**13,584**	**10,037**	**23,621**
	57.51%	**42.49%**	

The RA methodology was observed to make the number of compliant subjects—the congruous ones—slightly higher than the non-compliant subjects—the incongruous ones. The RA methodology suggests a higher level of compliance, and this higher level is equally present in all three classes in almost the same proportions.

This initial representation helps show some differences between the Istat and RA methodologies. First, group G1A is now not completely represented by

compliant subjects. It is worth recalling that the Istat methodology suggests that these subjects are already earning an income and a pension, and that other jobs can be considered a simple integration of safe and regular earnings. Because the revenues of another occupation are considered income from a secondary partial job, they are not investigated further. However, the RA methodology, which provides estimates of presumptive income for all taxpayers, also provides an estimation for this group, and presumptive incomes are then included in this analysis.

The predictions of the presumptive incomes of the investigated taxpayers can be summarized considering the Istat classification already provided:

Table 2.20: Presumptive operational incomes across groups.

	Declared	**Correction**	**Presumptive**
G1A	13,777	4,488	17,512
G1B	13,668	6,094	19,589
G1C	18,124	5,861	23,910
Total G1	**14,186**	**6,053**	**20,072**

In the previous table, the total for Group G1 represents the weighted means of the declared and presumptive incomes of all subjects included in the sample and the relative corrections. This information shows that the most numerous group, G1B, is also the group characterized by the highest average correction.

Presenting this same information for the other periods is also possible. If the proportion of subjects assumed to be non-compliant is investigated, approximately 70% of taxpayers assumed to be congruous (10% more considering the 2007 results) given the RA predictions.

A review of the incomes shows that the results are very close to each other, although the numbers change over the years. The same conclusions drawn for

Table 2.21: Distribution of congruous subjects among years – RA methodology.

Year	Congr.	%	Incongr.	%	Tot
2009	17,615	70.10%	7,515	29.90%	25,130
2010	17,268	67.93%	8,154	32.07%	25,422
2011	17,466	68.82%	7,915	31.18%	25,381

2007 can also be extended to the other years, including when the number of taxpayers changes:

Table 2.22: Declared income, correction, and presumptive income across years, RA methodology.

Year	n.	Declared	Correction	Presumptive
2007	23,621	14,186	6,053	20,072
2009	25,130	14,289	5,994	20,201
2010	25,422	14,676	4,519	19,052
2011	25,381	15,260	4,078	19,194

Compared with the Istat results, the values between the observed years are more similar. The correction ratio for the other years shows lower values for all years relative to the Istat years. Moreover, they are consistent for each year, and variations among the years are quite limited.

Table 2.23: Ratio of correction Group 1, RA methodology.

Year	Declared	Corrected	%
2009	359,081,559	507,661,835	141.38%
2010	373,081,476	484,345,553	129.82%
2011	387,303,089	487,154,145	125.78%

2.B.9 Comparison of methodologies

In the previous part, two different methodologies were illustrated and estimations were made on a specific database to show the different predictions by

these methodologies. This section investigates in greater detail the differences and examines the benefits and challenges of the possible effects of presumptive taxation.

As explained, the methodologies developed by Istat and RA are differentiated provided the diversities in the data and the intention of their investigations. Although the main goal is to provide the most correct estimation possible of real incomes, the results can be quite different.

The different methodologies can be compared by comparing the Istat and RA predictions. As previously noted, the analysis is for 2007, and investigations will determine whether the same results are replicated in other years.

Table 2.24: Distribution of subjects among classes.

	ISTAT	**RA**		**ISTAT**	**RA**	
	Compl.	Congr	Not Congr	**Non-compl.**	Congr	Not Congr
G1A	125	125	78			
G1B	6,635	4,636	1,999	14,044	7,102	6,942
G1C	284	236	48	2,455	1,485	970
Σ **G1**	7,044	4,997	2,125	16,499	8,587	7,912
		70.94%	30.17%		52.05%	47.95%

The two methodologies appear to be quite different in their predictions. The differences can be analyzed group by group. The initial Istat assumptions about G1A appear not to hold when the RA methodology is applied. If the RA methodology is implemented, only 40% of the sample is not congruous and, thus, is suspected of being characterized by tax evasion.

The analysis of Group G1B also shows that the Istat and RA predictions are inconsistent with each other. The prediction of compliancy and congruity are inconsistent; moreover, congruous taxpayers appear to be higher in number than incongruous taxpayers when Istat predicts non-compliance. The higher

proportion of congruous taxpayers relative to non-congruous taxpayers is only said to prevail when Istat predicts the subject to be compliant.

Finally, the same conclusions drawn for G1B are confirmed for G1C. The Istat predictions are inconsistent with the RA predictions, although congruence between the proportion of congruous and non-congruous taxpayers for this group appears to occur when Istat predicts that the subjects are compliant.

If the same analysis is performed on the other periods, the results are quite similar to the results of previous periods.

Table 2.25: Distribution of subjects among different classes in different years.

	ISTAT	RA		ISTAT	RA	
Year	Compl.	Congr	Not Congr	Not compl.	Congr	Not Congr
2009	6,827	5,230	1,657	18,243	12,385	5,858
		76.61%	24.27%		67.89%	32.11%
2010	7,484	5,879	1,605	17,938	11,389	6,549
		78.55%	21.45%		63.49%	36.51%
2011	7,121	5,651	1,470	18,260	11,815	6,445
		79.36%	20.64%		64.70%	35.30%

Compared with 2007, the results for the other periods appear consistent. When Istat predicts that a subject is compliant, the RA methodology prediction is that approximately two out of ten subjects are compliant. However, when Istat predicts non-compliance, the results are less consistent because six out of ten taxpayers are now supposed to be compliant according to the RA methodology.

In summary, both methodologies predict low tax compliance, with Istat predicting higher tax evasion. Interestingly, the methodologies' predictions are quite different in their identification of taxpayers as compliant or not compliant. This diversity can be justified by the differences in the methodologies that were previously illustrated, such as different starting assumptions and identifi-

cation strategies.

Bibliography

Abadie, A., D. Drukker, J. L. Herr, G. W. Imbens, et al. (2004). Implementing matching estimators for average treatment effects in stata. *Stata journal 4*, 290–311.

Advani, A., W. Elming, and J. Shaw (2015). How long-lasting are the effects of audits. *Tax Administration Research Centre*, 011–15.

Akerlof, G. A. (1978). The economics of" tagging" as applied to the optimal income tax, welfare programs, and manpower planning. *The American economic review 68*(1), 8–19.

Allingham, M. G. and A. Sandmo (1972). Income tax evasion: A theoretical analysis. *Journal of public economics 1*(3-4), 323–338.

Alm, J. (2012). Measuring, explaining, and controlling tax evasion: lessons from theory, experiments, and field studies. *International Tax and Public Finance 19*(1), 54–77.

Alm, J., G. H. McClelland, and W. D. Schulze (1992). Why do people pay taxes? *Journal of public Economics 48*(1), 21–38.

Alm, J. and B. Torgler (2006). Culture differences and tax morale in the united states and in europe. *Journal of economic psychology 27*(2), 224–246.

Andreoni, J., B. Erard, and J. Feinstein (1998). Tax compliance. *Journal of economic literature 36*(2), 818–860.

Balestrino, A. and U. Galmarini (2005). On the redistributive properties of presumptive taxation.

Beer, S., M. Kasper, E. Kirchler, and B. Erard (2015, 12). Audit impact study.

Boadway, R. and P. Pestieau (2006). Tagging and redistributive taxation. *Annales d'Économie et de Statistique*, 123–147.

Braiotta, A., A. Carfora, R. V. Pansini, and S. Pisani (2015). Tax gap and redistributive aspects across italy. *Argomenti di Discussion dellâ^{TM}Agenzia Delle Entrate*, 4–27.

Bulutoglu, K. (1995). Presumptive taxation. *Tax Policy Handbook*, 258–262.

Caliendo, M. and S. Kopeinig (2008). Some practical guidance for the implementation of propensity score matching. *Journal of economic surveys 22*(1), 31–72.

Cannari, L. and G. D'Alessio (2007). The opinion of italians on tax evasion.

Dal Prato, L. e. a. ((2016)). *"Studi di Settore"*.

DeBacker, J., B. T. Heim, A. Tran, and A. Yuskavage (2015a). Legal enforcement and corporate behavior: An analysis of tax aggressiveness after an audit. *The Journal of Law and Economics 58*(2), 291–324.

DeBacker, J., B. T. Heim, A. Tran, and A. Yuskavage (2015b). Once bitten, twice shy? the lasting impact of irs audits on individual tax reporting. *Journal of Financial Economics 117*(1), 122–138.

Fiorio, C. V., S. M. Iacus, and A. Santoro (2013). Taxpaying response of small firms to an increased probability of audit: some evidence from italy.

Franz, A. (1985). Estimates of the hidden economy in austria on the basis of official statistics. *Review of Income and Wealth 31*(4), 325–336.

Gamarra Rondinel, A. (2017). Evasion vs. real production responses to taxation among firms: bunching evidence from argentina.

Guala, F. and L. Mittone (2005). Experiments in economics: External validity and the robustness of phenomena. *Journal of Economic Methodology 12*(4), 495–515.

Hashimzade, N., G. D. Myles, F. Page, and M. D. Rablen (2014). Social networks and occupational choice: The endogenous formation of attitudes and beliefs about tax compliance. *Journal of Economic Psychology 40*, 134–146.

Immonen, R., R. Kanbur, M. Keen, and M. Tuomala (1998). Tagging and taxing: The optimal use of categorical and income information in designing tax/transfer schemes. *Economica 65*(258), 179–192.

Kakwani, N. C. (1977). Applications of lorenz curves in economic analysis. *Econometrica: Journal of the Econometric Society*, 719–727.

Kastlunger, B., E. Kirchler, L. Mittone, and J. Pitters (2009). Sequences of audits, tax compliance, and taxpaying strategies. *Journal of Economic Psychology 30*(3), 405–418.

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), 651–692.

Kleven, H. J., C. T. Kreiner, and E. Saez (2009). Why can modern governments tax so much? an agency model of firms as fiscal intermediaries. Technical report, National Bureau of Economic Research.

Larin, G. N. and M. N. Jacques (1994). *Is the Alternative Minimum Tax a Paper Tiger?* Faculté d'administration, Université de Sherbrooke.

Logue, K. D. and G. G. Vettori (2011). Narrowing the tax gap through presumptive taxation. *Colum. J. Tax L. 2*, 100.

Luttmer, E. F. and M. Singhal (2014). Tax morale. *The Journal of Economic Perspectives 28*(4), 149–168.

Maciejovsky, B., E. Kirchler, and H. Schwarzenberger (2007). Misperception of chance and loss repair: On the dynamics of tax compliance. *Journal of Economic Psychology 28*(6), 678–691.

Mazzolini, G., L. Pagani, and A. Santoro (2017). The deterrence effect of real-world operational tax audits. *Forthcoming*.

Mittone, L. (2006). Dynamic behaviour in tax evasion: An experimental approach. *The Journal of Socio-Economics 35*(5), 813–835.

Oecd (2015). *Tax Administration 2015. Comparative Information about Oecd and Other Advanced and Emerging Economies*. Oecd Publishing, Paris.

Oecd (2016). Italy's tax administration. Oecd Publishing, Paris.

Pashev, K. V. (2006). Presumptive taxation: lessons from bulgaria. *Post-Communist Economies 18*(4), 399–418.

Pomeranz, D. (2013). No taxation without information: Deterrence and self-enforcement in the value added tax. Technical report, National Bureau of Economic Research.

Puggioni, A. ((2015)). L'economia non osservata nei conti nazionali.

Salanie, B. (2002). *The Economics of Taxation*. Economica. Originally published in France under the title "Théorie économique de la fiscalité".

Santoro, A. and C. V. Fiorio (2011). Taxpayer behavior when audit rules are known: Evidence from italy. *Public Finance Review 39*(1), 103–123.

Shackelford, D. A., J. Slemrod, and J. M. Sallee (2007). A unifying model of how the tax system and generally accepted accounting principles affect corporate behavior. Technical report, National Bureau of Economic Research.

Slemrod, J. (2004). The economics of corporate tax selfishness. Technical report, National Bureau of Economic Research.

Slemrod, J. (2007). Cheating ourselves: The economics of tax evasion. *The journal of economic perspectives 21*(1), 25–48.

Slemrod, J. (2016). Tax compliance and enforcement: New research and its policy implications. *Available at SSRN*.

Slemrod, J. and S. Yitzhaki (1994). Analyzing the standard deduction as a presumptive tax. *International Tax and Public Finance 1*(1), 25–34.

Stotsky, J. (1995). Minimum taxes. *Tax Policy Handbook*.

Wallace, S. et al. (2002). Imputed an presumptive taxes: International experiences and lessons for russia. Technical report, International Center for Public Policy, Andrew Young School of Policy Studies, Georgia State University.

Yitzhaki, S. (2007). Cost-benefit analysis of presumptive taxation. *FinanzArchiv: Public Finance Analysis 63*(3), 311–326.