1. Introduction

Estimating the level and distribution of tax evasion is a difficult task. One traditional, aggregate method consists of a comparison between the income levels declared to fiscal authorities by taxpayers and estimates of taxable income obtained by processing national accounts. A more recent approach uses information derived from individual income surveys instead of national accounts. The underlying hypothesis is that some people may report their incomes more truthfully to a survey interviewer than on an income tax form. Even though the former may be underreported as well, the difference between the two sources should give an initial insight into the relative distribution of tax evasion. Usually, owing to the difficulty of obtaining individual fiscal data, this type of analysis is performed at an intermediate disaggregation level and the comparisons are carried out between the average income levels computed on microdata of different taxpayers belonging to similar groups, identified by income or demographic characteristics.

This paper was presented at workshop «Metodi quantitativi per l’analisi delle condizioni di vita: nuove concettualizzazioni, stime statistiche e procedure operative» held on 31st January 2009 at the University of Modena and Reggio Emilia. The workshop was organized within the project «Osservatorio sulle disuguaglianze sociali» of Fondazione Ermanno Gorrieri (www.fondazionegorrieri.it) jointly with CAPP (Centro di Analisi delle Politiche Pubbliche – University of Modena and Reggio Emilia) and CRIDIRE (Centro di Ricerca Interdipartimentale sulla Distribuzione del Reddito «C. Dagum» – University of Siena).

The authors are grateful to Giuliano Orlandi, director of the municipal statistical office of Modena, for his assistance in some steps of the survey process and to Giovanni Bigi who carried out the survey sample and linked the fiscal and the survey databases. The study was performed under the National Research Project «Costruzione di un’indagine sulle famiglie e di un modello di microsimulazione per l’analisi delle politiche sociali e fiscali a livello locale», financed by MIUR (Ministry of Education, University, and Research, year 2001, prot. 2001135524).
In this paper, to our knowledge for the first time for the Italian economy, the comparison between income declared to fiscal authorities – hereinafter Fiscal Income (FI) – and income reported in a survey – hereinafter Survey Income (SI) – is performed at the individual level, as the information on individual taxable income was derived from administrative sources, respecting anonymity, through exact matching with the sample of individuals interviewed in the survey on the economic conditions of households living in the Province of Modena in 2002 (ICESMO, *Indagine sulle Condizioni Economiche e Sociali delle famiglie della provincia di Modena*). Owing to the limited size of the geographical area, our results cannot be generalised to the whole country, but they do hold some interest in suggesting insights for further studies.

This research confirms the traditional results that the self-employed tend to evade more than employees, but it also shows a surprising result with reference to the correlation between the level (both absolute and relative) of tax evasion and income, which in our data appears to be positive.

The paper is organised as follows. Section 2 discusses some problems related to measurement errors in survey data and briefly reviews the main results of major studies on tax evasion in Italy. Section 3 describes the main characteristics of ICESMO and of fiscal data, and the basic features of the final sample obtained by merging the data survey file with the fiscal and census databases. Section 4 illustrates the measures of misreporting and tax evasion. Section 5 analyses the data and presents an econometric investigation of the determinants of relative tax evasion. Section 6 concludes.

2. *Estimating tax evasion on microdata*

The comparison between interviews and fiscal income reports is one of the ways of attempting to estimate tax evasion. The fundamental, though doubtful – see, *inter alia*, Moore *et al.* (1997) and Jäntti (2004) –, assumption of these studies is that income reported in surveys is more truthful than the income declared to tax authorities and the difference represents a proxy of tax evasion. A number of studies have been performed in Italy under this assumption, starting from Cannari *et al.* (1995), who compare the data collected by the SHIW of 1989 with declared tax incomes from administrative sources.

---

1 Studies based on exact matching are rare (see Flevotomou and Matsaganis, 2007). A good example is Jäntti (2004) for Finland.
Marenzi (1996) estimated the distribution profile of tax evasion by comparing the income in the SHIW of 1993 with the income declared on the tax forms for the year 1991. The frequency distribution of taxpayers, by classes of total income, was obtained from the Ministry of Finance tax return data. Tax evasion, measured by the difference between SI and FI, was assessed for four types of taxpayers: employees, pensioners, professionals and the self-employed and entrepreneurs. For employees the analysis was restricted to those who had filed a tax return (see below for details). A highly predictable result emerged from the comparison: tax evasion was higher among the self-employed than among employees. More puzzling was the evidence of significant evasion among wealthier pensioners. By income level, the tax evasion that appeared to be concentrated in the first two deciles for workers probably originated from undeclared jobs. For professionals, tax evasion was high in the low deciles: it was equal to 40% in the third decile and only 4% in the tenth. For entrepreneurs or the self-employed, tax evasion increased by value but was almost constant in percentage terms through the deciles (see also Bernasconi and Marenzi, 1997).

More recently, Fiorio and D’Amuri (2005) analysed tax evasion using the SHIW sample of 2000 and a random sample (250,000 taxpayers) selected and processed by the SECIT (Tax Advisory and Supervisory Service) of the Ministry of Finance, which includes all tax file sources (for employees, tax returns submitted by taxpayers, as well as data provided to the tax authorities on their employees by employers). The supposed tax evasion was analysed for two groups of taxpayers (employees and the self-employed), subdividing income into deciles. Their analysis confirmed the finding of higher evasion on the part of the self-employed. By income level, evasion seemed to be higher in the low deciles. One rather striking finding was negative evasion for the highest deciles of employees, appearing to show that rich taxpayers declare to the tax authorities more income than they actually earn. A possible explanation was the presence of measurement errors in the SHIW, an indication of the troublesome interaction between estimates of underreporting and evasion. The findings of Fiorio and D’Amuri (2005) imply that, since evasion is higher for low-income taxpayers, its presence makes the distribution of declared income more unequal than that of real income. Mantovani and Nienadowska (2007), however, raise the point that this conclusion postulates the absence of reranking between income distribution with and without tax evasion. Using data from the Bank of Italy survey and from the Ministry of Finance, they produce some evidence that reranking is actually present.

A more recent paper (Marino and Zizza, 2008) estimated IRPEF (income) tax evasion following the same lines as the preceding studies, by comparing 2004 SHIW data with tax returns filed for the same year, with a more
accurate procedure for harmonising the two different data sets. This allowed
a more detailed analysis by employment status, sex, and age. The result con-
firmed evasion among the self-employed and found lower tax compliance for
men and young taxpayers.

Though path-breaking and insightful, all these studies, in spite of their
use of microdata, suffer from the limitation of making comparisons between
grouped data, referred to different individual units. Our work, on the other
hand, represents the first attempt in Italy to compare SI e FI for the same
sample of individuals, with exact matching between the two incomes. Our
analysis is performed with reference to four employment status groups. The
proxy of tax evasion was the difference between the overall gross income de-
rived from ICESMO (SI) and from tax forms filed with the tax authorities
(FI).

3. Data sources

3.1. The survey data

The survey on the economic and social conditions of households living in
the Province of Modena (ICESMO) was carried out in 2002 and was based
on two-stage cluster sampling, stratified with respect to the area subdivided
in accordance with the socio-medical districts, in which municipalities were
the primary and the households the secondary sampling units. In this pa-
per we make use of only the subsample of 589 units resident in the town
of Modena. The interview questionnaire asked for information about a large
number of variables and, specifically, all those concerning the net amount of
earnings, real and personal estate, capital and financial assets referring to the
last twelve months. The collection period was concentrated between June
2002 and October 2002 with a tail up to December 2002. Therefore, the be-
ginning of the reference income period lasted from June 2001 to December
2001. This variation in reference income period between subjects is the most
critical aspect for at least two reasons. First of all, complex hypothetical op-
erations were necessary to obtain the individual SI comparable with the FI,
implying that the findings illustrated below might be the result of the mis-
match in the period rather than subjects’ mismatched behaviours. However,
this asymmetry is the same for all subjects, so it should not distort our re-
sults. Secondly, there is some evidence that a longer recall period tends to
lead to faulty recall (Moore et al., 1997).
3.2. Matching fiscal and survey data

Fiscal data were obtained from the statistics office of the City Council of Modena, which selected the ICESMO sampling units, four years after the selection. In fact, before drawing sample and immediately after data collection and processing, it was only possible to examine taxpayers individually through a system of interchange between the local tax register offices (SIA-TEL: Sistema di Interscambio dell’Anagrafe Tributaria degli Enti Locali) and the Ministry of Finance. The fiscal database was matched with the census database of 2001. The resulting file was matched with all the sampling units included in the survey design, obtaining four different sub-samples: refusals, untraceable units, reserves, and participant units who were the object of study. The types of tax forms (TF) utilized for producing the FI were TF-730, TF-Unico, and TF-770. The TF-730 and TF-Unico, produced by the taxpayer on an either-or basis (the former is more simplified and utilized by the majority of employees), contains all details of income sources (land, buildings, estate, employment, other, total, and taxable income) and tax deductions or tax allowances (for wife, children, relatives, employment, pension, and others) relevant for the application of personal income tax (IRPEF). The TF-770, which reports employers’ income, also contains information on the taxable income of employees, who are exempted from filing tax returns (as all the relevant information is provided by the employer).

The matching of files was satisfactory, but some errors were found. For example, information on gender was missing for 84 individuals in the fiscal database out of 1,388 (about 6%). The FI was missing for 357 individuals (25.7%), but for 254 individuals (18.3%) both SI and FI were missing because they were not obliged to file an income tax return. The possible errors fell in the secondary diagonal of Table 1 with a total amount equal to 10.4%, of whom 3.2% were total underreporters and 7.6% were total tax evaders or overreporters. The number of individuals with both SI and FI totalled 71.1%. If only SI is available, FI may be missing for two reasons (apart from

<table>
<thead>
<tr>
<th>FI= M. FI ≠ M.</th>
<th>SI ≠ Missing</th>
<th>SI= Missing</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>n/%</td>
<td>y</td>
<td>SD</td>
<td>n</td>
</tr>
<tr>
<td>FI ≠ Missing</td>
<td>987</td>
<td>21,807.2</td>
<td>21,496.2</td>
</tr>
<tr>
<td>SI</td>
<td>71.1%</td>
<td>20,751.4</td>
<td>20,607.5</td>
</tr>
<tr>
<td>SI ≠ Missing</td>
<td>106</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>SI</td>
<td>7.6%</td>
<td>13,436.2</td>
<td>14,056.1</td>
</tr>
<tr>
<td>Total n</td>
<td>1,093</td>
<td>295</td>
<td>1,388</td>
</tr>
</tbody>
</table>
errors in the matching process): either total evasion or exemption from the obligation to file a tax form, due to low income levels. In the analysis that follows, we also make use of the 106 observations with only SI information.

3.3. Summarised statistics of the sample

Summarised statistics of FI and SI of the restricted sub-sample (987 observations) are provided in Table 2, disaggregated by the type of tax form and gender. In particular, the income of women proved to be lower than that of men and the means of the two incomes were found to differ with regard to the type of tax form submitted. They also showed a significant interaction between tax form and gender. However, the differences between the SI and FI were not significant. The type of tax form was related to economic and employment status; in fact, the Unico showed lower coefficients than other types because it was used mainly by managers, top executives, atypical workers, professionals, the self-employed and entrepreneurs. They were individuals who did not like being interviewed about their income and did not trust interviewers asking about confidential information about property, stocks or income.

Figure 1 illustrates the distributions of FI and SI for 987 cases, where values over 100,000 € were eliminated to obtain greater readability and the histograms were smoothed using the Epanechnikov kernel function. It can be seen that the spike of SI (solid line) is higher than that of FI (dashed line) and at a lower level of income than that of FI, denoting the systematic underreporting of income amounts by respondents.
4. Some methodological problems on the measurement of tax evasion, misreporting and the unit of analysis

In accordance with the literature surveyed in the preceding section, our analysis assumes that $E_t$, the difference between SI and FI, is a good proxy of relative tax evasion. In fact, $E_t$ depends not only on tax evasion, but also on a number of causal factors that are very difficult to measure and disentangle, the most problematic being the difficulty in discriminating between evasion and income misreporting. The relationship among these two concepts and our analysis may be explained by the diagram below, where tax evasion, $E$, is indicated as the difference between the overall income liable to personal income tax (usually tax base gross of tax deductions), $F_t$, and the income declared to the tax authority, FI. Our work tries to identify $E$ starting from a non-fiscal notion of income, i.e. the income declared to the ICESMO interviewer, $S_t$, which diverges from $F_t$ for reasons which derive from the difference between the research aims of the survey and the measurement of personal income tax liability (tax erosion, lump-sum measurement of some incomes and costs, etc.). All these factors will be classified as objective misreporting, $U_o$. Moreover, «true» income SI may differ from $S_t$ because of the interview subject’s difficulty in recalling the relevant information accurately and, in general, due to cognitive problems dealt with in section 2.1. These factors are indicated in the diagram as subjective misreporting, $U_s$. The absolute difference between the two incomes, $E_t$ = SI – FI, measures $E$ – $U$, where $U$ = $U_o$ + $U_s$, which includes both subjective and objective misreporting (Figure 2).

The disentanglement of the measurement of $E$ from that of $U$ is impossible in the data from our sample. We can, however, formulate a priori
conjectures, where the degree of misreporting may be higher or lower. For instance, it seems reasonable to expect that more educated and wealthier people are better able to provide a correct assessment of their incomes. The evaluation of incomes derived from real estate may be more uncertain than those types of income that are ascertained on the basis of testable accounting procedures, like profits, or documented by wage receipts or tax forms. These aspects, though non-measurable, should be borne in mind in the interpretation of the results.

Income tax evasion may be computed in *absolute* or *relative* terms. The choice between these two measurements is not trivial, as it depends on the aims of the investigation. If one considers, for instance, that the resources that have been evaded could have been used for the production of public services whose benefits are shared in a uniform way by all citizens, it would be more appropriate to use the absolute difference between SI and FI as the indicator of the intensity of tax evasion. Very often, however, estimated tax evasion is evaluated as a percentage of the supposed true income: \( E,\% = 100 \frac{SI - FI}{SI} \), also referred to as percentage difference. In this case the indicator could, though not unambiguously, be interpreted as a *proxy* of an implicit ethical evaluation of an illegal behaviour. Existing studies provide both measures, but usually more emphasis is given to the relative indicator in the interpretation of the results. We will present the two measures as well, though we will rely more on the absolute difference, which seems to us more relevant for policy purposes.

A third methodological problem concerns the *choice of the unit of analysis*. Usually the literature makes reference to the taxpayer, whose main activ-
ity determines his or her employment status. However, it is obvious that a taxpayer may have income from different sources (labour, interest, rents, dividends, etc.), where the degree of temptation to evade tax differs. In principle, it would be possible to disaggregate all the various incomes and proceed on this basis to make a detailed analysis by income types. This would require data collection procedures of a complexity out of all proportion to the aims and limits of this research. Our analysis, like others along these same lines, therefore takes the taxpayer as a unit, and the prevalent form of income is used to identify his/her employment status.

5. The determinants of tax evasion

5.1. A descriptive analysis

We can now examine the data of the 1,093 individuals without missing SI (first column of Table 1), starting as usual from the relationship between estimated tax evasion, \( E_s = SI - FI \), and the employment status of the taxpayer. Then, we shall illustrate the income gradient of tax evasion. With reference to employment status, subdivided into three categories, Table 3 provides average incomes from tax files and from the survey, and the absolute and relative measures of tax evasion. Survey income is on average only marginally (+2.6%) higher than income from tax returns. This does not necessarily prove the existence of tax evasion, due to the possible presence, already discussed, of income misreporting in the survey. One can note, however, very significant differences among taxpayers of different employment status. Independent workers (a category that includes the self-employed, entrepreneurs and professionals) show a large gap between SI and FI (10,600 €, 28% of SI), non-existent for employees (–170 euro) and non-negligible, albeit negative, for pensioners and other taxpayers. In terms of differences in the propensity to evade by employment status, our data therefore seem to confirm the results of previous studies as far as the self-employed are concerned, while a different evaluation emerges for pensioners.

The results of the analysis of the relationship between tax evasion and income levels are more complex. Before starting the analysis, we must however discuss which income measure is to be compared with tax evasion. The most standard choice seems to be the individual income reported in the survey, SI. This is also consistent with the individual nature of the tax base for personal income taxation in Italy. One could, however, argue that the propensity to evade is influenced not so much by individual income as by the overall economic status and environment of the household, the place where consum-
tion and moral habits are formed, including the attitude to the moral obligation to pay taxes. In this case, household equivalent income could be a better choice.

A further problem that deserves attention is the treatment of extreme values. As we have shown in the preceding section, the estimated evasion shows high values of standard deviation, owing to the presence of some outliers, in particular in the right tail of the income distribution. In order to evaluate the significance of these observations for the overall results, it seems useful also to examine a trimmed sample: we therefore performed the analysis both on the whole sample and on the sample without the highest 1% and lowest 1% of the tax file values. Figures 3, 4, and 5 provide an initial illustration of these analyses.

Figure 3 shows the absolute and relative (i.e., as a percentage of survey income) levels of concealed income, by deciles of household equivalent disposable income, both for the whole sample and for the sample that rules out the top and bottom 1% of individuals, ordered by increasing values of absolute evasion. The two series of numbers are always very close, implying that the results are unaffected by the presence of exceptional values. In absolute terms, tax evasion is very low for most of the deciles of equivalent household income; it is actually concentrated in the tenth decile, while substantial underreporting is present in the first one. In the central deciles, $E_i$ is not particularly significant. Dropping the top and bottom 1% of observations with extreme values of evaded income does not significantly alter the picture. In relative terms, tax evasion is very high for the tenth decile, while in the bottom of the income distribution, problems of survey under-reporting seem to be significant.

Figures 4 and 5 report absolute tax evasion by deciles of individual survey income, first for the whole sample and then excluding individuals with zero fiscal income. In this case as well, evasion is strong in the upper deciles and generalised underreporting is present in the first three deciles. We can thus infer that zero fiscal incomes are relevant mainly for the lower deciles (probably owing to the presence of underground income and/or poor individuals below the taxable income threshold).

<table>
<thead>
<tr>
<th>Employment status</th>
<th>Fiscal income</th>
<th>Survey income</th>
<th>$E_i$</th>
<th>$E_i%$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Employee</td>
<td>21,101</td>
<td>20,961</td>
<td>-140</td>
<td>-0.7</td>
</tr>
<tr>
<td>Independent worker</td>
<td>27,710</td>
<td>38,311</td>
<td>10,601</td>
<td>27.7</td>
</tr>
<tr>
<td>Pensioners and others</td>
<td>15,020</td>
<td>13,259</td>
<td>-1,761</td>
<td>-13.3</td>
</tr>
<tr>
<td>Total</td>
<td>19,438</td>
<td>19,957</td>
<td>519</td>
<td>2.6</td>
</tr>
</tbody>
</table>

**Tab. 3. Absolute and relative tax evasion by employment status**
Considering the absolute measure of tax evasion by deciles of individual income, which is more comparable with other studies, our data offer a different picture of the distribution of tax evasion, which suggests the usefulness of a deeper analysis, crossing employment status and income levels. This is shown in Figure 5, which presents absolute tax evasion for employees, the self-employed and pensioners (this third category also includes other minor cases). What emerges is that tax evasion is fundamentally a problem of the self-employed and, though present at all income levels, it is particularly marked for the highest earners.

The literature surveyed in section 2 has investigated the relationship between estimated tax evasion and income distribution. A higher level of tax evasion for low-income taxpayers could be seen as a shelter against the worsening in the primary income distribution. Our data seem to weaken this argument which, as already said, in any case depends on the absence of rerank-
ing effects, which are however difficult to single out using grouped data. In our case, assuming that the difference between SI and FI is only attributable to tax evasion, exact matching allows us to compare the degree of mobility among deciles of taxpayers. Table 4 is a mobility matrix that shows the probability of moving to the different deciles of the FI distribution, starting from any given decile of SI; excluding the richest SI decile, for all the other quantiles most households move to a different FI decile, so that it is fair to conclude that the relevance of reranking seems to be high.

To complete our descriptive analysis, we consider the relationship between tax evasion and taxpayers’ level of education and age (Table 5). The results are summarised in the tables below. Though not uniformly, tax evasion is greater among taxpayers with high education levels and the non-elderly, with a peak at an age of around 50 years. A high level of education, age, and high levels of income do not appear to be personal characteristics positively associated with acceptance of a moral obligation to pay taxes.
An econometric analysis of the determinants of tax evasion

An analysis of the determinants of tax evasion may be a useful completion to our work. Among the possible results, here we present two estimates, where a measure of tax evasion is regressed on a set of personal characteristics: gender, age, income, employment status and level of education. The first equation estimates the determinants of absolute tax evasion, $E_t$. In the specification, the dummy for employment status interacts with the individual income level. The second estimate is a probit model where the dependent variable is one if the person is a tax evader (i.e., $p=1$ if $E_t>0$). Both equations confirm the preceding descriptive analysis, in particular the positive relationship between self-employment and income level. The estimates add the information that tax evasion seems to be higher among female taxpayers and significant for middle-aged taxpayers.

### Tab. 4. Mobility matrix between deciles of SI and FI

<table>
<thead>
<tr>
<th>Deciles of survey income</th>
<th>Deciles of fiscal income</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>1</td>
<td>29</td>
<td>48</td>
</tr>
<tr>
<td>2</td>
<td>14</td>
<td>16</td>
</tr>
<tr>
<td>3</td>
<td>8</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>8</td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td>19</td>
<td>3</td>
</tr>
<tr>
<td>6</td>
<td>7</td>
<td>5</td>
</tr>
<tr>
<td>7</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>8</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>9</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>10</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>Total</td>
<td>11</td>
<td>9</td>
</tr>
</tbody>
</table>

### Tab. 5. Declared tax income, survey income, absolute and relative tax evasion by education level and age of taxpayers

<table>
<thead>
<tr>
<th>Education level</th>
<th>Fiscal income: FI</th>
<th>Survey income: SI</th>
<th>$E_t = S - F$</th>
<th>$E_t%$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elementary</td>
<td>12,840</td>
<td>12,075</td>
<td>765</td>
<td>-6.3</td>
</tr>
<tr>
<td>Intermediate</td>
<td>15,950</td>
<td>16,181</td>
<td>231</td>
<td>1.4</td>
</tr>
<tr>
<td>High school</td>
<td>20,791</td>
<td>22,048</td>
<td>1,257</td>
<td>5.7</td>
</tr>
<tr>
<td>University degree</td>
<td>31,485</td>
<td>32,976</td>
<td>1,491</td>
<td>4.5</td>
</tr>
<tr>
<td>Total</td>
<td>19,438</td>
<td>19,957</td>
<td>519</td>
<td>2.6</td>
</tr>
<tr>
<td>Age classes</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤ 40</td>
<td>18,149</td>
<td>19,860</td>
<td>1,710</td>
<td>8.6</td>
</tr>
<tr>
<td>41-50</td>
<td>21,527</td>
<td>22,615</td>
<td>1,088</td>
<td>4.8</td>
</tr>
<tr>
<td>51-60</td>
<td>20,512</td>
<td>24,999</td>
<td>4,486</td>
<td>17.9</td>
</tr>
<tr>
<td>≥ 61</td>
<td>19,207</td>
<td>16,740</td>
<td>-2,467</td>
<td>-14.7</td>
</tr>
<tr>
<td>Total</td>
<td>19,438</td>
<td>19,957</td>
<td>519</td>
<td>2.6</td>
</tr>
</tbody>
</table>
Conclusions

The value-added of this paper lies in the use of individual microdata, obtained by exact matching between administrative and survey data, for the investigation of absolute and relative tax evasion. As usual in the literature, we adopt the difference between survey and tax declared income as a proxy of tax evasion. After discussing some methodological problems, in particular the impossibility of disentangling tax evasion from misreporting, we performed an empirical analysis of the relationship between tax evasion and a set of personal characteristics of taxpayers.

Our results support the widely shared belief that in Italy the propensity to evade is higher among the self-employed. At the same time, our data do not support the existence of a negative correlation between tax evasion and income level, found in the literature, which in this work appears to be positive. Underreporting seems to be concentrated in the lowest deciles, and tax evasion is higher among the self-employed.

\[ \begin{array}{ccc}
\text{Men} & \text{Coefficient} & \text{Standard error} \\
41-50 & 3,579.442 & 1,744.052 \\
51-60 & 1,049.103 & 1,908.097 \\
\geq 61 & 781.5988 & 1,835.915 \\
\text{Intermediate} & 8,231.878 & 2,006.399 \\
\text{High school} & 4,994.612 & 1,814.663 \\
\text{University degree} & 4,192.158 & 1,545.173 \\
Y^{\text{self-employed}} & 0.452789 & 0.034381 \\
Y^{\text{employee}} & 0.183596 & 0.037624 \\
\text{Costant} & -8,828.659 & 1,871.242 \\
\end{array} \]

\[ z \]

\[ \begin{array}{ccc}
\text{Men} & \text{Coefficient} & \text{Standard error} \\
41-50 & 0.3832787 & 0.1554946 \\
51-60 & 0.0350757 & 0.1670281 \\
\geq 61 & 0.2592578 & 0.1377442 \\
\text{Intermediate} & 0.0591589 & 0.1396367 \\
\text{High school} & -0.0263730 & 0.1239535 \\
\text{University degree} & 0.1181641 & 0.1064125 \\
Y^{\text{self-employed}} & -0.0084901 & 0.0137910 \\
Y^{\text{employee}} & 0.0478666 & 0.0150934 \\
\text{Costant} & -0.3832348 & 0.1314320 \\
\end{array} \]

Note: Number of observation = 1,093; R-squared = 0.1485.

Note: Number of observation = 1,093; Pseudo R-squared = 0.0293.

6. Conclusions

The value-added of this paper lies in the use of individual microdata, obtained by exact matching between administrative and survey data, for the investigation of absolute and relative tax evasion. As usual in the literature, we adopt the difference between survey and tax declared income as a proxy of tax evasion. After discussing some methodological problems, in particular the impossibility of disentangling tax evasion from misreporting, we performed an empirical analysis of the relationship between tax evasion and a set of personal characteristics of taxpayers.

Our results support the widely shared belief that in Italy the propensity to evade is higher among the self-employed. At the same time, our data do not support the existence of a negative correlation between tax evasion and income level, found in the literature, which in this work appears to be positive. Underreporting seems to be concentrated in the lowest deciles, and tax evasion is higher among the self-employed.
evasion in the highest two. In our sample women tend to underreport more than men and underreporting and tax evasion show a dependence on age.

Notwithstanding the more reliable statistical base, the results must be accepted with caution, as the measure of tax evasion adopted is not able to clearly distinguish between evasion and misreporting, a problem that can be overcome only by investing in the production of more detailed survey data.

References


**Tax evasion and misreporting in income tax returns and household income surveys**

by Massimo Baldini, Paolo Bosi and Michele Lalla

*Summary*: This paper studies tax evasion and misreporting by comparing data from income tax files and a survey of household incomes, both relative to a medium-size Italian town. The novelty of our approach lies in the use of individual microdata, obtained by *exact matching* between the two data sources. As usual, the adopted proxy for tax evasion is the difference between survey income and tax-declared income. Our results lend support to the
widely shared belief that in Italy the propensity to evade is higher among the self-employed. However, our findings do not bear out the existence, which emerged in previous studies, of a negative relationship between tax evasion and income levels. Underreporting seems to be concentrated in the lowest income deciles, while the level of tax evasion is higher at the top of the income distribution. Notwithstanding the reliability of the statistical base, our results must be accepted with caution, as the measure of tax evasion adopted is not able to make a clear distinction between evasion and misreporting.

Keywords: income underreporting, income tax evasion, survey income, administrative data.

J.E.L. Classification: C46; C81; D31; H26; Z13.

Address:
Massimo Baldini, Dipartimento di Economia politica, Università di Modena e Reggio Emilia, via Berengario 51, I-41100 Modena. E-mail: massimo.baldini@uniomre.it
Paolo Bosi, Dipartimento di Economia politica, Università di Modena e Reggio Emilia, via Berengario 51, I-41100 Modena. E-mail: paolo.bosi@uniomre.it
Michele Lalla, Dipartimento di Economia politica, Università di Modena e Reggio Emilia, via Berengario 51, I-41100 Modena. E-mail: michele.lalla@uniomre.it