# Tax evasion in Italy:

## an analysis using a tax-benefit microsimulation model*

Carlo V. Fiorio†

University of Milan and Econpubblica

Francesco D'Amuri‡

University of Pavia and Econpubblica

**Abstract**

In this paper we use a direct method to estimate tax evasion in Italy assuming that tax evaders might consider declaring a closer-to-true income in an anonymous interview. The methodology is applied to work income only, as pension income cannot be hidden to tax authorities and capital income is measured with large error in available survey data sets. The data sets considered are the 1998 and 2000 Survey of Household Income (SHIW) by the Bank of Italy and the 1998 and 2000 tax forms table produced by SeCIT. Posing particular attention to the post-stratification of the data, we find that tax evasion is consistently higher for self-employment income than for employment income and it is larger at bottom deciles, although some under-sampling problems need to be considered. The pattern of work income concealment found shows that personal income tax evasion reduces the average tax rate but it also increases the progressivity of the tax system. This result is however driven by the large values of income avoidance found in bottom deciles, which might itself be due to the under-sampling of income receivers with poorest income. The results are consistent across the two years considered.

# 1 Introduction

Tax evasion in Italy is a serious issue: between a quarter and half of the GDP seems to be hidden to the tax authorities (Schneider, 2000a). These aggregate figures

*We are grateful to Alberto Zanardi, Anna Marenzi and Massimo Bordignon for important suggestions and to SeCIT for providing the tax forms tables.

†Corresponding author. Address: Department of Economics, Business and Statistics, University of Milan, Via Conservatorio, 7, 20122 Milan, Italy. email: carlo.fiorio@unimi.it.

‡Address: Department of Public Economics, University of Pavia, Strada Nuova, 65, 27100 Pavia, Italy. email: francesco.damuri@unibocconi.it.

are of great importance at a macro level, for instance for the reliability of official statistics and the efficiency of national productions. However they do not provide insights to policy makers that wish to investigate who are tax evaders and to start understanding the reasons why some taxpayers might consider under-declaring their income.

Measuring tax evasion is all but simple: Schneider (2000b, p. 1) describes tax evasion measurement "as a scientific passion for knowing the unknown". However, tax evasion analysis is relevant for public policy design and for estimating the bias that tax evasion introduces in some statistics, both at the macro and at the micro level. Section 2 provides a brief review of recent contributions on the estimation of tax evasion in Italy, focussing mainly on microeconomic approaches and in particular on direct methods of tax evasion estimation. Section 3 critically analyzes the representativeness of the data set and describes a grossing-up procedure to correct major deviations from population totals. This procedure is then implemented in the empirical analysis of tax evasion presented in Section 4. Section 5 explains how tax evasion is estimated and introduced in a tax-benefit microsimulation model, Section 6 analyzes redistributive effects of work tax evasion applying a microsimulation model for Italy and Section 7 concludes.

## 2 Available evidence about tax evasion in Italy

Tax evasion can be estimated using a direct or an indirect approach. Indirect methods estimate tax evasion considering it equal to the difference between aggregated macro indicators (e.g the discrepancy between income and expenditures or the difference between the actual demand for money and the demand for money estimated in absence of taxes). Direct methods aim at estimating tax evasion through the use of sample survey micro-data based on voluntary participation or the results of the auditing activity of tax authorities. In contrast to indirect methods, direct methods are more suitable to analyze tax evasion at the micro level and they can point out directions for policy.

Some of these methods have been applied to provide a measure of tax evasion in Italy. Among those who used indirect methods, Schneider (2000a) used the currency demand approach, Zizza (2002) also the factorial analysis. Zizza estimates the share of the underground economy (excluding illegal and criminal activities) on GDP for the years 1984-2000 between a maximum of 17.6% (1991) and a minimum of 14.3% (2000). Schneider's estimates include also illegal and criminal activities. According to him the share of the underground economy on the italian GDP is very high and increasing (from 25.8% in 1994 to 27.8% in 1998), the highest rate among the OECD countries.

Calzaroni (2000), Bernasconi and Marenzi (1997), Marenzi (1996), and Cannari et al. (1995) used direct methods. Calzaroni (2000) estimates labor supply and labor demand functions by sectors using household and firm surveys, respectively, and compares results at the national and the regional level. The difference between the two is considered to be the number of the irregular workers. This figure, multiplied for the average sectorial productivity estimated for regular workers gives a first measure of the underground economy. The overall incidence of the underground economy is calculated complementing this figure with coefficients correcting for the underestimation of the turnover and the balancing between aggregated input and output. The results indicate that, for 1998, the share of the underground economy on GDP in Italy is between 14.7 and 15.4%. Cannari et al. (1995), Bernasconi and Marenzi (1997) and Marenzi (1996) use a different approach: they assume that individuals report a more truthful income to an anonymous interviewer than to the tax authorities. Hence, when income data recorded in the Survey of Household Income and Wealth (SHIW) - produced by the Bank of Italy (BI) - is larger than that recorded in the analysis of tax forms - produced by the Ministry of Finance (MF) - the difference between BI and MF disposable income is considered as hidden income. Cannari et al. (1995) and Marenzi (1996) considered years 1989 and 1991, respectively. The estimate of tax evasion is performed for different groups of tax payers, identified by their main income (employment, self-employment, pension, etc.). Marenzi (1996) finds evidence of positive tax evasion in the two first deciles of employment income and negative evasion in the following ones, an evasion increasing in the level of income for pensioners and a tax evasion that is decreasing, in relative terms, in the level of income for entrepreneurs and professionals. In Cannari et al. (1995) tax evasion is on average zero for total employment income (for this kind of taxpayers an estimate by decile is not provided) and decreasing in the level of income for members of the art or professions and for entrepreneurs. Bernasconi and Marenzi (1997) and Marenzi (1996) provide also an estimate of redistributive effects using a tax-benefit microsimulation model. They showed that had evasion not incurred, the vertical effect of taxation would have increased and the horizontal and re-ranking effect would have decreased by large proportions.

This methodology relies heavily on the hypothesis that the SHIW data set is representative of the population and of its subgroups. As MF data refers to the population, a measure of tax evasion based on this methodology requires the BI data to be a good approximation of the population. This requirement must be verified carefully.

An estimate of tax evasion stemming form the comparison between BI and MF data also requires that income variables are defined consistently. Cannari et al. (1995), Bernasconi and Marenzi (1997) and Marenzi (1996) use total net income by group of taxpayers. However, the BI data set is quite reliable for the measurement of work income but it is much less so for other types of income such as capital, estate and building income (Cannari and D'Alessio, 1992). This is due to two main reasons: first, these data are collected at the household level and they can only be imputed to the individual taxpayer; second, there is a tendency to misestimate the true value of these incomes, which is probably not voluntary and however common also to other similar surveys. For these reasons, we suggest here to focus only on work incomes.

A final point refers to the imputation of tax evasion and the analysis of redistributive impacts of tax evasion using a microsimulation model. Bernasconi and Marenzi (1997) seem to assume that evaded income is a uniform percentage of disposable income for individuals belonging to the same income decile. However, this procedure is likely to cause biases in the distribution of imputed incomes as evasion rates can be very different among contiguous deciles.

## 3 The issue of grossing-up

Before applying a direct methodology for tax evasion estimation based on the comparison between a survey-based data set and tax forms, it is important to analyze the non-response bias. The non-response bias is larger the larger is the rate of non-response and the larger are the difference between respondents and non-respondents (see among others Little and Rubin, 1987). Unfortunately the non-response rate in the SHIW is rather high (Table 1) and some studies show that the non-response decision is not random. Cannari and D'Alessio (1992) analyzed the non-response bias in the SHIW using the second wave of the panel sub-sample (the first wave was interviewed in 1987, the second in 1989). Knowing the characteristics of those who refused to respond again in 1989, they expanded the results to the whole sample . They found that non-response is more frequent in households who reside in urban areas and in the North. The participation rates decline as income rises and household size decreases, while the relationship with the age of the head of the households is ambiguous. D'Alessio and Faiella (2002) also showed that non-response behavior is dependent on net financial wealth. They found this result using a supplementary sample of about 2,000 households, clients of a leading commercial bank. Although the sample can hardly be considered representative of the whole population and sub-samples size are not very large, individuals with financial wealth larger than Lit 1 billion (about €0.5 million) have about half the response rates of other groups. D'Alessio and Faiella (2002) used also alternative methodologies reaching the conclusion that non-response behavior is not random, and is more frequent among wealthier households. This implies that the post-stratification techniques traditionally employed on a few known demographic characteristics of the population, such as sex and age, cannot fully account for the non-response bias.

In this paper we care about making the characteristics of the individual in the

| year | response rate |
|---|---|
| 1987 | 64.3% |
| 1989 | 38.4% |
| 1991 | 33.2% |
| 1993 | 57.8% |
| 1995 | 56.9% |
| 1998 | 43.9% |
| 2000 | 38.3% |

**Table 1:** Sources: Brandolini (1999) and Banca d'Italia (2002)

SHIW data set as close as possible to those of the population. The procedure of grossing-up is concerned with generating figures to cover the population being modelled from the data set under use. The procedure should adjust for differences between the sample data and the characteristics of the population to be modelled at the date of sampling. The grossing-up procedure is basically aimed at adjusting the data set to reflect differential non-response between different groups in the sample. It involves stratifying the sample, after the data have been collected, by some relevant characteristics, and applying known proportions. This procedure is also sometimes referred to as post-stratification (see for instance Atkinson and Micklewright, 1983). The grossing-up procedure consists in assigning to each unit in a sample of dimension $N$ a weight $p_j$, with $j = 1, ..., N$, such that some chosen statistics of interest calculated on the weighted sample coincide with the population statistics. The procedure is trivial if we want to reconcile the sample with the population using only one discrete statistic, $s_k$ with $k = 1, ..., K$, such as family types or income ranges. In this case, we compute the probability of having the characteristic $s_k$ in the sample, say $P(s_k)$, and make it equal to the probability of having the same characteristic in the population, say $p(s_k)$. If the dimension of the sample and of the population are $N$ and $n$ respectively, then the grossing-up weight is $p_j = np(s_k)/NP(s_k)$, i.e. the size of the cell with characteristic $s_k$ in the population divided by the size of the cell with characteristic $s_k$ in the sample. If more variables are included for the grossing-up procedure, the interactions between the different variables (i.e. the their joint distribution) should be considered. However, this conflicts with available information from external sources, which in general do not report the joint distribution of population variables but only the totals for each variable. For instance, it is generally possible to know the total number of single-parent families and the total number of self-employed in the population but not how many single-parent families have self-employment income. Hence, the conditions imposed on the weights $p_j$ are far less stringent than in the "full information" case we would have if the joint distributions were known, and in general there are many possible sets of weights $p_j$ achieving the desired adjustment. To choose among them Atkinson et al. (1988) suggest the requirement that, given a data set of dimension $N$, with original sampling weights $q_j$, $j = 1, 2, ..., N$, the set of grossing-up weights $p_j$ have the least deviation from original weights, $q_j$. The original weights could reflect the sampling procedure or be uniform. Both grossing-up and initial weights have to sum up to the population size: $\sum q_j = \sum p_j = n$. If original and grossing-up weights sum up to the sample dimension, they first have to be multiplied by $n/N$. It is then common practice to impose the condition that the new weights minimize the distance from initial weights. In order to avoid negative weights, Atkinson et al. (1988) suggest minimizing a measure of distance derived from information theory (see for instance Cowell, 1980):

$$d(p,q) = \sum p_j log\left(\frac{p_j}{q_j}\right)$$

As for the optimal number of control totals to be included, no result is currently available. Although it is more common to face the problem of not having enough external sources than of having too many, Sutherland (1989, p. 15) warns on the risk of increasing the variance of weights since the larger the number of control totals becomes, the smaller the number of observations in each "cell" (i.e. with each combination of characteristics being controlled for). Moreover, a particular set of grossing-up weights can be able to closely reflect the characteristics of the population as for some variables but not for others.

The SHIW data set is post-stratified using the variables sex, age class, area and dimension of the town of residence Banca d'Italia (2000, p. 40). However, it is not clearly stated what methodology was used and, for instance, which age classes were considered. Table 2 and 3 show how much does the weighted sample differ from population totals using the grossing-up weight provided by the BI in SHIW 1998 and SHIW 2000, respectively. It can be seen that, using the BI weights, the differences between the grossed-up and actual figures are small (less than 1%) as far as sex and area of residence (North-West (NW), North-East (NE), Center (C) and South (S)) are considered, but it becomes worryingly large for age groups, occupation (especially by area of residence) and schooling. It should also be noted that the number of self-employed people (excluding owners or members of a family business and active shareholders and partners) are under-represented by over 20% in 1998 and over-represented by a similar proportion in 2000 with respect to the population of tax payers. This shows a problem with grossed-up simulations: for instance the effects of an hypothetical tax policy that affected mainly self-employed or the young in the South in 1998 would probably be underestimated as these groups are under-represented using the corresponding BI weights. These distortions could be even worse for other sub-samples, which we do not consider here. All these issues are of relevance whenever an analysis of income by population sub-groups is performed. For these reasons a set of alternative grossing-up weights were estimated using the same methodology as Atkinson et al. (1988) using control totals found in Ministero delle Finanze (2002, 2004) and ISTAT (2004). As we focus our attention on work income we mainly used the "Weight 5", which matches total number of employed and self-employed taxpayer with population totals, besides controlling for a number of other variables, such as distribution of age groups, and level of education.

# 4 Estimating tax evasion

We use a direct approach to tax evasion estimation, as previously used for Italy by Marenzi (1996) and Cannari et al. (1995). The basic assumption is that an income receiver who decides to evade tax payment will under-report her taxable income to tax authorities but declare the true income, or at least a closer approximation to the true income, to an interviewer who grants anonymity. As survey-based data tend

| Variable | External sources* | BI | | Our Weight | |
|---|---|---|---|---|---|
| | Totals (a) | Totals (b) | Diff (b/a-1) | Totals (c) | Diff (c/a-1) |
| Total population | 57,612,615 | 57,612,568 | 0.0% | 57,612,767 | 0.0% |
| Males | 27,967,670 | 27,951,136 | -0.1% | 27,967,726 | 0.0% |
| Females | 29,644,945 | 29,661,432 | 0.1% | 29,645,041 | 0.0% |
| Pop NW | 15,069,493 | 15,099,744 | 0.2% | 15,069,553 | 0.0% |
| Pop NE | 10,560,820 | 10,547,936 | -0.1% | 10,560,830 | 0.0% |
| Pop C | 11,071,715 | 11,064,505 | -0.1% | 11,071,779 | 0.0% |
| Pop S | 20,910,587 | 20,900,383 | 0.0% | 20,910,605 | 0.0% |
| Age<=18 | 10,845,419 | 11,032,994 | 1.7% | 10,845,461 | 0.0% |
| 18<Age<=30 | 9,987,651 | 9,619,324 | -3.7% | 9,987,666 | 0.0% |
| 30<Age<=65 | 27,218,646 | 26,787,452 | -1.6% | 27,218,724 | 0.0% |
| Age>65 | 9,560,899 | 10,172,798 | 6.4% | 9,560,916 | 0.0% |
| Age<=18 NW | 2,409,663 | 2,497,552 | 3.6% | 2,409,673 | 0.0% |
| Age<=18 NE | 1,687,699 | 1,853,786 | 9.8% | 1,687,700 | 0.0% |
| Age<=18 C | 1,873,809 | 2,073,762 | 10.7% | 1,873,820 | 0.0% |
| Age<=18 S | 4,874,248 | 4,607,894 | -5.5% | 4,874,268 | 0.0% |
| 18<Age<=30 NW | 2,498,184 | 2,411,373 | -3.5% | 2,498,203 | 0.0% |
| 18<Age<=30 NE | 1,766,221 | 1,630,855 | -7.7% | 1,766,226 | 0.0% |
| 18<Age<=30 C | 1,824,075 | 1,988,102 | 9.0% | 1,824,081 | 0.0% |
| 18<Age<=30 S | 3,899,171 | 3,588,994 | -8.0% | 3,899,156 | 0.0% |
| 30<Age<=65 NW | 7,509,728 | 7,523,230 | 0.2% | 7,509,771 | 0.0% |
| 30<Age<=65 NE | 5,174,474 | 5,070,504 | -2.0% | 5,174,468 | 0.0% |
| 30<Age<=65 C | 5,368,887 | 5,191,858 | -3.3% | 5,368,923 | 0.0% |
| 30<Age<=65 S | 9,165,557 | 9,001,860 | -1.8% | 9,165,562 | 0.0% |
| Age>65 NW | 2,651,918 | 2,667,589 | 0.6% | 2,651,906 | 0.0% |
| Age>65 NE | 1,932,426 | 1,992,791 | 3.1% | 1,932,436 | 0.0% |
| Age>65 C | 2,004,944 | 1,810,783 | -9.7% | 2,004,955 | 0.0% |
| Age>65 S | 2,971,611 | 3,701,635 | 24.6% | 2,971,619 | 0.0% |
| Employed | 14,549,000 | 14,530,169 | -0.1% | 14,549,043 | 0.0% |
| Employed NW | 4,470,000 | 4,345,113 | -2.8% | 4,470,034 | 0.0% |
| Employed NE | 3,104,000 | 3,199,310 | 3.1% | 3,103,995 | 0.0% |
| Employed C | 2,911,000 | 2,821,364 | -3.1% | 2,911,019 | 0.0% |
| Employed S | 4,064,000 | 4,164,382 | 2.5% | 4,063,995 | 0.0% |
| Self-employed | 5,886,000 | 5,852,953 | -0.6% | 5,886,013 | 0.0% |
| Self-employed NW | 1,643,000 | 1,793,760 | 9.2% | 1,643,003 | 0.0% |
| Self-employed NE | 1,330,000 | 1,156,244 | -13.1% | 1,329,998 | 0.0% |
| Self-employed C | 1,184,000 | 1,532,649 | 29.4% | 1,184,009 | 0.0% |
| Self-employed S | 1,729,000 | 1,370,300 | -20.7% | 1,729,003 | 0.0% |
| Elementary schooling | 16,104,000 | 19,785,184 | 22.9% | 16,104,039 | 0.0% |
| Compulsory schooling | 16,118,000 | 15,489,547 | -3.9% | 16,118,080 | 0.0% |
| High school degree | 13,365,000 | 15,415,524 | 15.3% | 13,365,006 | 0.0% |
| Laurea | 3,066,000 | 3,641,053 | 18.8% | 3,066,008 | 0.0% |
| Agriculture | 1,201,000 | 1,038,245 | -13.6% | 1,200,999 | 0.0% |
| Industry | 6,730,000 | 6,548,547 | -2.7% | 6,730,019 | 0.0% |
| Services | 12,504,000 | 12,796,330 | 2.3% | 12,504,038 | 0.0% |

*Source: Our calculations on SHIW98*

* External sources from ISTAT

**Table 2:** Discrepancies between population and weighted SHIW sample. Year 1998

| Variable | External sources* | BI | | Our Weight | |
|---|---|---|---|---|---|
| | Totals (a) | Totals (b) | Diff (b/a-1) | Totals (c) | Diff (c/a-1) |
| Total population | 57,844,017 | 57,828,424 | 0.0% | 57,844,122 | 0.0% |
| Males | 27,796,000 | 28,068,065 | 1.0% | 27,796,040 | 0.0% |
| Females | 30,048,017 | 29,760,359 | -1.0% | 30,048,082 | 0.0% |
| Pop NW | 15,153,050 | 15,151,408 | 0.0% | 15,153,116 | 0.0% |
| Pop NE | 10,681,233 | 10,645,864 | -0.3% | 10,681,238 | 0.0% |
| Pop C | 11,159,583 | 11,108,163 | -0.5% | 11,159,581 | 0.0% |
| Pop S | 20,850,151 | 20,922,989 | 0.3% | 20,850,187 | 0.0% |
| Age<=19 | 11,349,415 | 11,495,187 | 1.3% | 11,349,399 | 0.0% |
| 19<Age<=65 | 35,938,667 | 35,962,509 | 0.1% | 35,938,761 | 0.0% |
| Age>65 | 10,555,935 | 10,370,728 | -1.8% | 10,555,962 | 0.0% |
| Age<=19 NW | 2,562,196 | 2,837,546 | 10.7% | 2,562,197 | 0.0% |
| Age<=19 NE | 1,814,818 | 1,949,195 | 7.4% | 1,814,818 | 0.0% |
| Age<=19 C | 1,983,300 | 2,150,925 | 8.5% | 1,983,291 | 0.0% |
| Age<=19 S | 4,989,101 | 4,557,521 | -8.7% | 4,989,093 | 0.0% |
| 19<Age<=65 NW | 9,654,836 | 9,644,351 | -0.1% | 9,654,898 | 0.0% |
| 19<Age<=65 NE | 6,752,727 | 6,805,086 | 0.8% | 6,752,741 | 0.0% |
| 19<Age<=65 C | 6,969,449 | 7,061,154 | 1.3% | 6,969,445 | 0.0% |
| 19<Age<=65 S | 12,561,655 | 12,451,918 | -0.9% | 12,561,677 | 0.0% |
| Age>65 NW | 2,936,018 | 2,669,511 | -9.1% | 2,936,021 | 0.0% |
| Age>65 NE | 2,113,688 | 1,891,583 | -10.5% | 2,113,679 | 0.0% |
| Age>65 C | 2,206,834 | 1,896,084 | -14.1% | 2,206,845 | 0.0% |
| Age>65 S | 3,299,395 | 3,913,550 | 18.6% | 3,299,417 | 0.0% |
| Employed | 15,131,000 | 15,051,997 | -0.5% | 15,131,026 | 0.0% |
| Employed NW | 4,616,000 | 4,464,179 | -3.3% | 4,616,032 | 0.0% |
| Employed NE | 3,247,000 | 3,302,600 | 1.7% | 3,247,003 | 0.0% |
| Employed C | 3,050,000 | 3,213,664 | 5.4% | 3,049,985 | 0.0% |
| Employed S | 4,218,000 | 4,071,554 | -3.5% | 4,218,006 | 0.0% |
| Self-employed | 5,949,000 | 5,867,783 | -1.4% | 5,949,004 | 0.0% |
| Self-employed NW | 1,678,000 | 1,598,946 | -4.7% | 1,678,012 | 0.0% |
| Self-employed NE | 1,367,000 | 1,297,700 | -5.1% | 1,366,997 | 0.0% |
| Self-employed C | 1,205,000 | 1,302,585 | 8.1% | 1,204,992 | 0.0% |
| Self-employed S | 1,699,000 | 1,668,552 | -1.8% | 1,699,003 | 0.0% |
| Elementary schooling | 19,766,000 | 19,628,313 | -0.7% | 19,766,086 | 0.0% |
| Compulsory schooling | 16,556,000 | 15,624,763 | -5.6% | 16,555,998 | 0.0% |
| High school degree | 14,291,000 | 15,436,445 | 8.0% | 14,291,011 | 0.0% |
| Laurea | 3,546,000 | 3,799,227 | 7.1% | 3,546,001 | 0.0% |
| Industry | 6,767,000 | 7,051,688 | 4.2% | 6,766,990 | 0.0% |
| Services | 13,193,000 | 12,362,652 | -6.3% | 13,193,026 | 0.0% |

*Source: Our calculations on SHIW00*

* External sources from ISTAT

**Table 3:** Discrepancies between population and weighted SHIW sample. Year 2000

to grant anonymity to increase the probability of participation in the survey and of truthful declarations, the comparison of income distribution using tax records and survey-based data sets allows one to have a picture of tax evasion behavior.

The difference between survey grossed-up income and population tax forms data can be considered as the sum of underground economy (tax avoidance and evasion concerning legal activities) and of informal economy (individual activity with low level of organization, based on individual and familiar relationship, such as baby sitting, domestic cleaning, etc.). Criminal or illegal economy (e.g. tax avoidance and evasion due to illegal activities such as drug trafficking or unauthorized medical practice) is not included since we believe that those incurring in such activities are very unlikely to accept the interview. For an analogous reason we believe that total tax evaders are not among respondents of this kind of surveys.

In this paper the 1998 and 2000 Survey of Household Income and Wealth dataset (SHIW), produced by the Bank of Italy (BI), is used. The SHIW data sets are then compared with the tables of the Ministry of Finance (MF) on incomes and tax returns referring to 1998 and 2000, respectively. The main difference between the two data sources is that the former is a sample survey, the latter presents population data. The SHIW collects detailed re-call information of income as well as individual characteristics of each component of the interviewed household, but these data are likely to be flawed by underreporting, miss-reporting and under-sampling, especially for some type of incomes, such as estate and building income or capital income. The MF data sets report only income and tax data about the population of taxpayers in tabulated form (e.g. by income groups, by area of residence, etc.): imputation of before- and after-tax income by occupation is feasible to the cost of some approximations.

The SHIW data collect information on disposable income only, and not on the amount of taxes paid, which can only be simulated: some additional degree of measurement error might come from recording mistakes of the interviewer, imprecise answers of the interviewed household who are not required to provide evidence of their incomes. MF data instead, come directly from tax forms and they should only include tax-payer mistakes in filling in tax form and tax form data elaboration mistakes.

Notwithstanding these limitations we believe that such a comparison is informative, though it presents wide margins of improvement provided better data are produced and made available. The exercise we perform here is similar to that of Marenzi (1996) and it can be seen as an update of that paper. However, our work differs from Marenzi (1996) for three main reasons: (a) we focus on employment and self-employment[1] income only, as they make the larger part of the personal income

[1]According to our definition, self-employed are members of the arts and or professions, sole propietors and free lances. Owners or members of a family business and active shareholders or partners are excluded.

tax base and as measurement error of other type of incomes (mainly capital, real estate and building incomes) is huge in the SHIW data sets (see Brandolini, 1999). We also disregard differences in pension income as they might reasonably come only from measurement error rather than from explicit underreporting behavior as pensions are paid by state institutions, which are also responsible to deduct the due tax before payment. (b) We carefully consider the issue of grossing-up. (c) We check for robustness of our conclusions applying the methodology to year 1998 and year 2000[2].

Results are shown in Tables 4 and 5. A first point to be noted is that our analysis is not based on exact deciles. An income distribution by deciles is obtained by ordering incomes in ascending order and by dividing them in ten groups with the same number of individuals. However, MF tables are not organized by deciles (i.e. groups with the same number of individuals) but by over 30 ascending income groups, each with different number of taxpayers included. Hence, we regrouped population taxpayer incomes in order to have a distribution as close as possible to a distribution by deciles[3]. We preferred not to impute deciles in the MF tables as we verified that small changes of deciles can greatly influence the estimated amount of hidden income. Since the distribution of incomes in the BI data set is nearly continuous, a decile distribution that closely matches the MF distribution was constructed using the SHIW data set. The percentage of tax evasion is then estimated as the ratio between the mean after-tax income within a SHIW decile and the mean after-tax income within the corresponding MF decile. Income tax in MF tables is imputed to different types of income depending on the relative importance of each gross income components.

Table 4 shows that employment income is underreported mainly in bottom deciles. According to our estimates, nearly 70% of the lowest 11% of employment incomes are not declared to the tax authorities in 1998. The concealed income declines as income increases and becomes about zero for incomes over the median.

[2]The BI and MF data sets used refer to the same year so that the bias due to indexation of incomes using some sort of price index is avoided.

[3]Although it would be more correct to refer to these as percentiles, for simplicity we will henceforth refer to them as deciles.

| Year 1998 | | | | |
|---|---|---|---|---|
| Employment income | | | | |
| Percentile | BI employment income | MF employment income | Evasion (euro) | Evasion (%) |
| 11 | 3073 | 980 | 2093 | 68.1% |
| 22 | 7087 | 4212 | 2876 | 40.6% |
| 30 | 9269 | 7196 | 2073 | 22.4% |
| 42 | 10723 | 9662 | 1062 | 9.9% |
| 57 | 12140 | 11723 | 417 | 3.4% |
| 69 | 13141 | 13446 | -305 | -2.3% |
| 79 | 14861 | 15094 | -233 | -1.6% |
| 89 | 17141 | 17286 | -145 | -0.8% |
| 93 | 19869 | 20419 | -550 | -2.8% |

| Year 2000 | | | | |
|---|---|---|---|---|
| Employment income | | | | |
| Percentile | BI employment income | MF employment income | Evasion (euro) | Evasion (%) |
| 11 | 1456 | 3175 | 1719 | 54.1% |
| 21 | 5269 | 7243 | 1974 | 27.3% |
| 31 | 8259 | 9928 | 1669 | 16.8% |
| 49 | 11444 | 11700 | 256 | 2.2% |
| 62 | 13681 | 13129 | -553 | -4.2% |
| 73 | 15384 | 14928 | -456 | -3.1% |
| 86 | 17590 | 17221 | -369 | -2.1% |
| 92 | 20646 | 19806 | -840 | -4.2% |

**Table 4:** Difference of mean employment income declared to the Bank of Italy (BI) and to the Ministry of Finance (MF)

| Year 1998 | | | | |
|---|---|---|---|---|
| Self employment income | | | | |
| Percentile | BI employment income | MF employment income | Evasion (euro) | Evasion (%) |
| 11 | 1741 | 1219 | 522 | 30.0% |
| 23 | 4262 | 4024 | 238 | 5.6% |
| 31 | 6256 | 5964 | 292 | 4.7% |
| 41 | 7973 | 7406 | 567 | 7.1% |
| 55 | 10266 | 9148 | 1118 | 10.9% |
| 61 | 12122 | 10484 | 1638 | 13.5% |
| 74 | 14495 | 12139 | 2355 | 16.2% |
| 86 | 18788 | 16561 | 2227 | 11.9% |
| 90 | 24433 | 21401 | 3032 | 12.4% |

| Year 2000 | | | | |
|---|---|---|---|---|
| Self employment income | | | | |
| Percentile | BI employment income | MF employment income | Evasion (euro) | Evasion (%) |
| 10 | 1185 | 2279 | 1095 | 48.0% |
| 21 | 4042 | 4865 | 823 | 16.9% |
| 32 | 6355 | 7206 | 851 | 11.8% |
| 43 | 8241 | 9876 | 1636 | 16.6% |
| 51 | 9733 | 11801 | 2069 | 17.5% |
| 64 | 11140 | 14506 | 3366 | 23.2% |
| 72 | 12967 | 17606 | 4639 | 26.4% |
| 85 | 16805 | 20658 | 3853 | 18.7% |
| 93 | 23887 | 27965 | 4078 | 14.6% |

**Table 5:** Difference of mean self-employment income declared to the Bank of Italy (BI) and to the Ministry of Finance (MF)

This pattern of tax evasion at lower level of incomes is probably due to the high frequency of irregular position at low income levels of employment incomes. The negative values of income concealment in larger percentiles is instead more likely to be due to the under-sampling of high incomes in SHIW (recall Section 2). An analogous, though slightly lower, pattern of tax evasion is found in year 2000, confirming the conclusion that employment income is more frequently underreported at low levels. The reduction of tax evasion at low levels of income could be due to the introduction of temporary and part-time contracts in the job market at the end of the 1990s.

Table 5 shows the pattern of underreporting of self-employment income. As for employment income, tax evasion is more likely at lower levels of income. However, the rate of tax evasion is consistently positive for higher incomes. The pattern of self-employment income tax evasion in 2000 is about the same than in 1998, though the percentage of evaded income tends to be slightly higher.

Figure 1 shows a smoothed picture of the amount of concealed income depending on type of income, in a considered year. More employment income is hidden than self-employment incomes at low levels of income. However, while average concealed self-employment incomes can be as high as €4,000 for higher incomes, it is about zero for employment income.

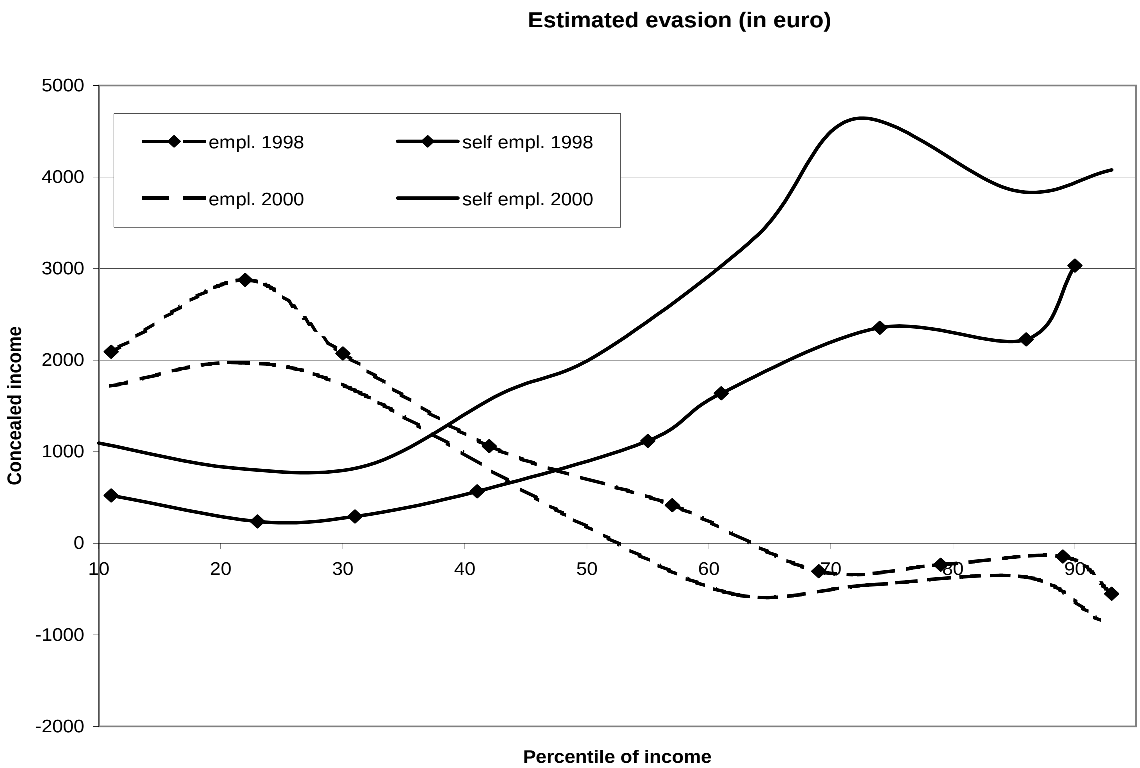


**Figure 1:** Estimate of amount of income concealed to tax authorities

# 5 Analysis of redistributive effects of tax evasion using a microsimulation model

The results presented in Section 4 rise a set of concerns about equity and progressivity of taxation. It was shown that probability of evading taxation is not evenly distributed across different income groups and across different types of incomes.

Bernasconi and Marenzi (1997) assume that people belonging to the same income decile and with equal type of income have the same compliance behavior. In other words, they suppose that there is no re-ranking and horizontal inequity within deciles but only across them. "It is evident that the difference between the two data sources (BI and MF), computed by decile, is interpretable as evasion of the decile only if it is assumed that evasion did not move taxpayers from a decile to another before and after tax evasion" (Bernasconi and Marenzi, 1997, p. 24). Hence they estimate the amount of evaded income comparing incomes by deciles in the two data sources. Although we agree that among restricted groups of taxpayers it is not unreasonable to assume homogeneity of tax compliance behavior, we find this imputation of tax evasion problematic. Given this pattern of tax evasion, we believe that its main limitation is not re-ranking but jumps it induces in the imputed income distribution. Applying Bernasconi and Marenzi (1997) methodology, a *discontinuous* piecewise linear evasion function would be obtained. For instance, let's consider Table 4: an employee with a true income equal to €5,165 declares to the tax authorities an income that is 68.1% smaller (€1,648). Someone with a true income that is just €1 more declares an income much larger (€3,068=€5,165(1-40.6%)). We believe that such a jump is unlikely to occur and difficult to justify.

Hence, we impute evaded income in a different way. A coefficient ($k_j$, $j = 1, ...q$, where $q$ is the number of income groups (deciles) considered) is computed such that evaded income function is continuous in income. A piecewise linear function is used, which closely resembles a typical tax-bracket function. In other words, we assume that the marginal evasion tax rate is constant and equal to $k_1$ for incomes below the first percentile, it changes to $k_2$ for incomes exceeding the first percentile but not the second one, and so on. Evasion rates $k_i$, $i = 1, ..., q$, which are different by year and occupation, are obtained by numerical approximation from Tables 4 and 5, holding the constraint that the SHIW mean income (deducted by the amount of imputed evasion) is equal to the MF mean income of the corresponding decile[4]. The

[4]Results can be replicated downloading the relevant files from

amount of evaded income using this continuous piecewise linear function is clearly different from what imputed by the discontinuous piece-wise linear one suggested by Bernasconi and Marenzi (1997) (see Figure 2).

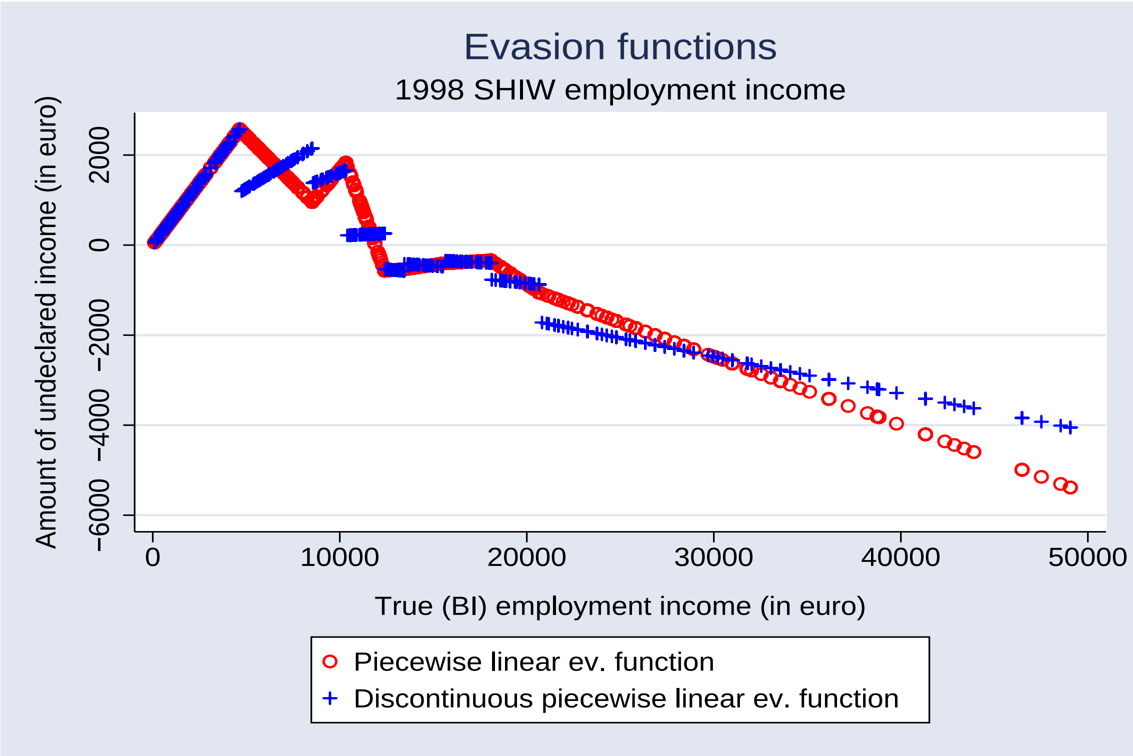


**Figure 2:** Evasion functions with different imputation hypotheses

The total amount of evaded income by individual $i$ ($E_i, i = 1, ..., N$) can be defined as follows. Let $Y_i$ be her disposable income as recorded in SHIW, $B_j$ be the $j$-th income decile, with $B_0 = 0$ and $j = 1, ...q$, $k_j$ and $q$ be the marginal evasion rate and the number of deciles, respectively, as defined above, then:

$$E_i = \sum_{j=1}^{q-1} Z_j(B_j - B_{j-1})k_j + I_j(Y_i - B_{j-1})k_{j+1} \tag{1}$$

where

$$I_j = \begin{cases} 1 & if \quad B_{j-1} < Y_i \leq B_j \\ 0 & if \quad\quad\quad o.w. \end{cases}$$

http://www.econpubblica.uni-bocconi.it/fiorio.

and

$$Z_j = \begin{cases} 1 & if \quad B_{j-1} < Y_i \\ 0 & if \qquad o.w. \end{cases}$$

# 6 Effects of tax evasion on incidence and progressivity indices

Results presented in Section 4 rise a set of concerns about incidence and progressivity of taxation: they show that the proportion of income that is hidden to the tax authorities is not evenly distributed across different income groups and across different types of incomes.

As in Bernasconi and Marenzi (1997) we assume that people belonging to the same income interval and with equal type of income have the same compliance behavior. In other words, we suppose that there is no re-ranking and horizontal iniquity within the same income interval but only across them. "It is evident that the difference between the two data sources (BI and MF), computed by decile, is interpretable as evasion of the decile only if it is assumed that evasion did not move taxpayers from a decile to another before and after tax evasion" (Bernasconi and Marenzi, 1997, p. 24).

As the SHIWs record only disposable income after all taxes and social contributions, a microsimulation model is used to simulate before-tax income as well as to perform simulations. The TABEITA98 and TABEITA00 models used are static microsimulation models that simulate personal taxation (IRPEF and "imposte sostitutive") using the SHIW98 and SHIW00, respectively[5]. In order to avoid overestimating the amount of tax paid and of before-tax income, only income net of (imputed) tax evasion is used to simulate before tax evasion. The imputation of tax evasion is obtained from tables 4 and 5, using the coefficient $k_i$, $i = 1, ..., 10$ as in (1). In deciles where income concealment is negative, it is set to zero.

Results of the simulation are analyzed using some indices for the measurement of the effects of taxation and the Gini coefficient, before and after taxes. The

[5] For details about the model, see Fiorio and D'Amuri (2004)

Kakwani index is a very popular index of progressivity: it measures the departure from proportionality as the difference between the concentration coefficient of tax $C_t$ and the Gini index of before-tax income, $G_y$:

$$K_t = C_t - G_y \tag{2}$$

For large samples the minimum value of the Kakwani index is $-(1 + G_y)$ and the maximum value is $1 - G_y$. The first case happens when the poorest person pays all the tax ($C_t = -1$), the second when all the tax is paid by the richest person, leading to maximal progressivity (Kakwani, 1977).

The redistributive effect looks at the shift from before-tax to after-tax income. With no re-ranking, the after-tax Lorenz curve coincides with the after-tax income concentration curve. The Reynolds-Smolensky index ($RS$) is equal to the difference between the Gini coefficient of before-tax income ($G_y$) and the concentration coefficient of after-tax income ($C_{y-t}$) (Reynolds and Smolensky, 1977). In absence of re-ranking it is the reduction of the Gini coefficient achieved by the tax. It is also equal to the product of a progressivity index (e.g. $K_t$) and the average tax on net income ($t/1-t$):

$$RS = G_y - C_{y-t} = \frac{t}{1-t} K_t \tag{3}$$

Hence the redistributive effect is determined by disproportionality and tax incidence. However, as the re-ranking effects are likely to occur with the tax system, the Reynolds-Smolonsky index, which is an indicator of vertical equity, should be written as the sum of a redistributive effect (RE) and a re-ranking effect (RR) (Lambert, 1993, p. 185):

$$RS = RE + RR = (G_y - C_{y-t} - G_{y-t} + C_{y-t}) + (G_{y-t} - C_{y-t}) \tag{4}$$

These indices allow us to measure the importance of our estimates of work tax evasion on the progressivity, redistribution and incidence of personal income taxation. Consistently with the rest of the paper we focussed on work income receivers only. Let us denote the evaded income - as estimated in Section 4 - with E, the before-tax income declared to fiscal authorities with BT, the net personal income tax applying the tax code on BT with PIT, the after-tax income with AT (AT = BT - PIT). Finally we denote the true BT (TBT = BT + E) with TBT, and the true AT (i.e. the disposable income after all PIT is paid on TBT) with TAT. Using the estimates of Section 4, two sets of incomes were then simulated: (a) the "*status quo*", where some conceal part of their work income; (b) the "*ideal world*", where no income is concealed. The following exercise was performed:

**The "*status quo*":**

**Step 1** The net income as declared in SHIW and deducted by E is fed into TABEITA00 to obtain BT;

**Step 2** Using TABEITA00, PIT and AT are obtained as an output of the microsimulation model.

**Step 3** TBT and TAT are then computed adding E to both BT and AT.

**The "*ideal world*":**

**Step 1** as in "*status quo*";

**Step 2** TBT is computed as the sum of BT and E.

**Step 3** TABEITA00 is used on TBT to compute TAT.

Step 1 in the exercise above is necessary as, by definition, no tax is paid on E and the SHIW00 data do not provide information about PIT paid. TBT and TAT are then used in (4) to analyze the effects of income concealment on personal income taxation.

Table 6 presents results. In the "*status quo*", the tax relative to net income in 1998 is equal to 24% (24.4% in 2000) and it would be larger by 1.7% (1.8% in 2000) if no income was concealed. The redistributive effect would however be slightly smaller in the "*ideal world*" as the departure from proportionality (the $K_t$

index) would be about 4.2% smaller (4% in 2000) in the "*ideal world*". This result - together with a (slightly) larger Gini index in the "*ideal world*" after tax-income - might be puzzling to some readers, who would expect evasion to increase inequities. However they should recall that the direct methodology applied in Section 4 found large evasion in lower deciles of employment income. Moreover these indices were computed using the subsample of those declaring some work income only: work tax evasion would certainly have a lower effect in case the whole sample was considered.

**1998**

| | RS | K | RR | t/(t-1) | Gini BT | Gini AT |
|---|---|---|---|---|---|---|
| **Ideal world** | 0.0544 | 0.2110 | 0.0002 | 0.2576 | 0.3888 | 0.3346 |
| **Status quo** | 0.0568 | 0.2359 | 0.0005 | 0.2407 | 0.3888 | 0.3325 |

**2000**

| | RS | K | RR | t/(t-1) | Gini BT | Gini AT |
|---|---|---|---|---|---|---|
| **Ideal world** | 0.0505 | 0.1932 | 0.0009 | 0.2612 | 0.3798 | 0.3302 |
| **Status quo** | 0.0526 | 0.2158 | 0.0017 | 0.2436 | 0.3798 | 0.3290 |

Source: Our calculations on SHIW98, SHIW00 using TABEITA98, TABEITA00.

**Table 6:** Indices of progressivity, redistribution and inequality

# 7 Concluding remarks

This paper provides an estimate of work tax evasion in Italy using a direct methodology. The key assumption is that an income receiver who decides to evade taxes will under-report her taxable income to tax authorities but declare the true income, or at least a closer approximation to it, during an anonymous interview. The analysis was performed on work income only, as other types of income are recorded with much greater approximations in available survey data. A great care was put on the grossing-up of the sample to population totals, as grossed-up statistics using weights included in SHIWs would bias the tax evasion estimation. Results show that employment income is evaded at low levels of income but evasion rate decreases constantly and becomes negligible after the median income. Self-employment income is instead evaded by a positive amount, regardless of the income range considered: as their self-employment evasion behavior is about constant across income levels, individual absolute amounts of concealed income are larger.

The robustness of results was checked replicating the same analysis on two different years (1998 and 2000): in both cases the pattern of employment and self-employment income does not change significantly. The redistributive analysis of tax evasion shows that, had evaded income been completely declared, inequality among income receivers would have increased as a larger share of income is hide away at lower deciles.

Some points about this methodology should however be noted: first, we only focussed on work income evasion to isolate possible misreporting of other type of incomes; second, the hypothesis adopted can only provide a conservative estimate of work tax evasion; third, this methodology rules out the possibility of estimating total evaders, as we assume that they do not accept to be interviewed; forth, a possible reason why work tax evasion increases progressivity of the tax system is likely to come from the fact that it is mainly localized at low income levels, which would present a very low effective tax rate if declared. Our results about hidden income are closer to those of Calzaroni (2000), who also focussed on work income, though using a different methodology.

Our results show that a policy to contrast tax evasion should focus mainly on low income employment and self-employment. However, while self-employed are likely to under-declare their income to reduce their personal income tax, employees' motivation are probably different as tax deductions and tax credits would allow poor employees to keep their effective tax rate close to zero. Poor employees might consider to conceal part of their income to cash in at least part of the amount of total social contributions or to maintain some means-tested social assistance (e.g. nursery in some municipalities). Policies to contrast tax evasion should keep into consideration this differences in tax evasion motivations.

# References


Atkinson, A. B., Gomulka, J., and Sutherland, H. (1988). Grossing-up FES data for Tax-Benefit Models. pages 223–253. Atkinson and Sutherland (1988), Tax-Benefit Models, STICERD Occasional Paper,No. 10, London. London School of Economics, STICERD.

Atkinson, A. B. and Micklewright, J. (1983). On the Reliability of Income Data in the Family Expenditure Survey 1970-1977. *Journal of the Royal Statistical Association*, 146(1):33–61.

Banca d'Italia (2000). *I bilanci delle famiglie italiane nell'anno 1998*, volume X of *Supplementi al Bollettino Statistico. Note metodologiche e informazionistatistiche.* Banca d'Italia, Roma.

Banca d'Italia (2002). *I bilanci delle famiglie italiane nell'anno 2000*, volume XII of *Supplementi al Bollettino Statistico. Note metodologiche e informazionistatistiche.* Banca d'Italia, Roma.

Bernasconi, M. and Marenzi, A. (1997). Gli effetti redistributivi dell'evasione fiscale in Italia. *Ricerche quantitative per la politica economica*, pages 1–38. Convegno CIDE-SADIBA, Banca d'Italia, novembre.

Brandolini, A. (1999). The distribution of personal income in post-war italy: Source description,data quality, and the time pattern of income inequality. *Giornale degli Economisti e Annali di Economia*, 58:183–239.

Calzaroni, M. (2000). L'occupazione come strumento per la stima esaustiva del PIL e la misuradel sommerso. contributo al seminario dell'ISTAT, Roma, 12-13, gennaio.

Cannari, L., Ceriani, V., and D'Alessio, G. (1995). *Il recupero degli imponibili sottratti a tassazione.* Ricerche quantitative per la politica economica, sadiba, banca ditalia, 3-5 novembre edition.

Cannari, L. and D'Alessio, G. (1992). Mancate interviste e distorsione degli stimatori. *Temi di discussione del Servizio Studi, Banca d'Italia*, (172).

Cowell, F. A. (1980). On the structure of additive inequality measures. *Review of Economic Studies*, 47:521–31.

D'Alessio, G. and Faiella, I. (2002). Non-response behavior in the Bank of Italy's Survey of Household Incomeand Wealth. *Temi di discussione del Servizio Studi*, (462). Banca d'Italia.

Fiorio, C. V. and D'Amuri, F. (2004). TABEITA: a tax-benefit microsimulation model for Italian households. mimeo.

ISTAT (2004). *Popolazione & statistiche demografiche*. ISTAT, Istituto nazionale di statistica, Rome. Data available at `http://demo.istat.it/`.

Kakwani, N. (1977). Measurement of tax progressivity: an international comparison. *Economic Journal*, 87:71–80.

Lambert, P. J. (1993). *The distribution and redistribution of income: a mathematical analysis*. Manchester University Press, Manchester, 2nd edition.

Little, R. J. A. and Rubin, D. B. (1987). *Statistical Analysis with Missing Data*. Wiley, New York.

Marenzi, A. (1996). Prime analisi sulla distribuzione dell'evasione dell'IRPEF per categoriedi contribuenti e per livelli di reddito. pages 305–341. in Rapporto CNEL, il Mulino.

Ministero delle Finanze (2002). *Analisi delle dichiarazioni dei redditi. Il modello Unico 1999/redditi 1998:le persone fisiche*. Ministero delle Finanze, Rome.

Ministero delle Finanze (2004). *Analisi delle dichiarazioni dei redditi. Il modello Unico 2001/redditi 2000:le persone fisiche*. Ministero delle Finanze, Rome.

Reynolds, M. and Smolensky, E. (1977). *Public expenditures, taxes and the distribution of income: the United States,1950, 1961, 1970.* Academic Press, New York.

Schneider, F. (2000a). The increase of the size of the shadow economy of 18 OECD countries: somepreliminary explanations. *IFO Working Paper*, (306).

Schneider, F. (2000b). The value added of underground activities: Size and measurement of the shadoweconomies and shadow economy labor force all over the world. *World Bank Working Paper*. Washington, D.C.

Sutherland, H. (1989). Constructing a Tax-Benefit Model: What Advice Can One Give? In *Taxation, incentive and the distribution of income*, number 141. STICERD, London School of Economics.

Zizza, R. (2002). Metodologie di stima dell'economia sommersa: un'applicazione al caso italiano. *Temi di discussione del Servizio Studi*, (463). Banca d'Italia.