Income under reporting and tax evasion in Italy

Estimates and distributive effects

DOCUMENTO DI VALUTAZIONE N. 8

DOCUMENTO
DI VALUTAZIONE

Ufficio Valutazione Impatto
Impact Assessment Office

Senato della Repubblica

Questo *Documento di valutazione* è a cura di

ANDREA ALBAREA

MICHELE BERNASCONI

ANNA MARENZI

DINO RIZZI

Università Ca' Foscari di Venezia

I dati sono aggiornati ad ottobre 2017

CODICI JEL: C20, C63, D31, H24, H26, H31

PAROLE CHIAVE: TAX EVASION, INCOME UNDERREPORTING, HOUSEHOLD SURVEYS, PERSONAL INCOME TAX, TAX-BENEFIT MICROSIMULATION

Income under reporting and tax evasion in Italy

Estimates and distributive effects

Gennaio 2018

Abstract

The paper estimates the extent of evasion of personal income tax (PIT) in Italy by integrating two methods that the literature has previously applied separately. The consumption-based method introduced by Pissarides and Weber (1989) is used to estimate underreporting of income in micro data collected in the household IT-SILC survey. We adopt an econometric specification close in spirit to that of Feldman and Slemrod (2007), which allows us to estimate income underreporting at different rates for different income sources. The underreporting estimates are then used in the discrepancy method to correct the incomes declared by the households in the survey and to compare them with administrative data. The comparison provides new estimates of evasion of personal income tax by type of income, region and income class. The estimates are used to improve microsimulation analyses of the distribution impact of tax evasion.

Summary

1. Introduction

Measuring tax evasion is often described as attempting to obtain "evidence on the invisible" (Slemrod and Weber 2012).[1] Several approaches have been developed to obtain evidence on tax evasion that depend on the purpose of the analysis and particularly on which effects of tax evasion one wants to measure.

In this paper we propose an approach that integrates two methods that the literature has previously applied separately. Both methods adopt a microeconomic perspective. The analysis focuses on the personal income tax (PIT) in Italy (Irpef—*"imposta sui redditi delle persone fisiche"*—and other local income taxes) and also studies the distributive effects of this type of tax evasion.

The first method, known as the consumption-based approach, was developed by Pissarides and Weber (1989). It uses micro-economic observations from consumption-expenditure surveys to estimate the consumption function for certain classes of goods. After controlling for several household characteristics, the method assumes differences in consumption propensities estimated for various categories of income earners as measures of the official tax-return income that is underreported by the various income categories. In particular, the method assumes that all categories report consumption expenditures accurately, while incomes are reported correctly by only some categories (reference categories) of income earners. Under the assumption that individuals report the same value level of income to the surveys as they do in their official tax returns, it follows that income underreporting in the surveys can be used to measure the evasion committed by the various income categories. For example, Pissarides and Weber (1989) used employees as the reference category, while the self-employed were estimated as substantially underreporting their official tax-return income. The method has since been applied to estimate underreporting rates in various countries and for other income categories. (Studies include Besim and Jenkins 2005, Feldman and Slemrod 2007, Engström and Holmlund 2009, Hurst et al. 2014, Ekici and Besim 2016, and several others quoted in section 2.) However, as far as we know, the method has never been applied to Italy.

One criticism of the consumption-based method lies in its assumption that people behave in the survey as they do in reporting their official tax-return income. In fact, a diverse hypothesis is behind a different micro-economic method to measure evasion. The alternative method is based on comparisons between the income distributions from the surveys and the income distribution derived from the official tax-return data. Typically, these kinds of comparisons show that the distributions obtained from the surveys suggest higher incomes than the distributions obtained from the official tax registers. After taking into account all possible explanations for the disparities between the two distributions (e.g., disparities that are due to official tax-return data's often referring to gross income,

[1] Slemrod and Weber (2012) also provided a methodological review of various approaches, distinguishing between methods that start from a micro standpoint—or, as they are sometimes called, bottom-up or direct methods—and approaches that are based on macro-economic aggregates, which are referred to as top-down or indirect methods. (See also Schneider 2005, Alm and Embaye 2013, and Schneider and Enste 2013.)

while survey data tend to refer to after-tax incomes), the remaining differences can be interpreted as measures of evasion. For this reason, the procedure is also called the discrepancy method.[2] Clearly, the method builds on the hypothesis that people report their income truthfully in surveys, just as they do variables like consumption and expenditures, because they trust that their data will not be disclosed to the tax authorities, eliminating the incentive to lie.

The discrepancy method is often combined with microsimulation analyses, and the evasion rates estimated by the discrepancy method are then often employed by tax-benefit microsimulation models to compare the actual income distributions to counterfactual distributions simulated assuming full tax compliance in order to determine the distribution impact of tax evasion. Analyses carried out with this approach have been conducted to investigate tax evasion in Italy (Marenzi 1996, Cannari et al. 1997, Bernasconi and Marenzi 1999, Fiorio and D'Amuri 2006, Marino and Zizza 2008, Baldini et al. 2009, Albarea et al. 2015) and other countries (e.g. Matsaganis et al. 2010, Figari et al. 2012).

Despite the intuition on which the discrepancy method is built, a large literature has identified various biases that affect people’s answers to surveys, among which one of the most important is the tendency to underreport income.[3]

The present study combines the consumption-based method and the discrepancy approach to estimate tax evasion in Italy. We start by considering the possibility of income underreporting in the Italian Survey of Income and Living Conditions (IT-SILC), which is the Italian part of the European Statistics on Income and Living Conditions. We take an econometric specification that combines Pissarides and Weber’s (1988) model with Feldman and Slemrod’s (2007) more recent approach in order to estimate income underreporting at different rates on different income sources. The model encompasses both the case in which a given income source is underreported at the same rate by all taxpayers and the case in which underreporting rates are affected by idiosyncratic propensities to evade. Then we use the underreporting estimates to correct the net incomes indicated by households in the IT-SILC survey and use the discrepancy method to compare the results with the official tax-return data to provide new estimates of PIT evasion in Italy. Finally, we study the distributive impact of tax evasion using the BETAMOD tax-benefit microsimulation model (Albarea et al. 2015).

We obtain several results. Previous studies that have applied the discrepancy method to micro data have often reported lower evasion rates than have studies that were conducted at the macro level (e.g. Marino and Zizza 2008, and section 3.3. for references). Including the underreporting in the discrepancy method improves the estimates’ alignment with macro studies, almost doubling to 14.4 percent the overall evasion rate estimated for PIT (about 7.25%). The econometric analysis confirms that self-employment incomes and rental incomes are substantially underreported in the IT-

[2] The discrepancy method can also be applied in macro studies to obtain aggregate measures of tax evasion by, for example, comparing statistics on aggregate income, aggregate expenditure, labour force, etc. (See Giovannini 2011 for a review of studies that have used both micro and macro based methods in Italy.)

[3] The large literature that has discussed the various types of measurements errors that can affect data sets that provide income information, including income measures obtained from surveys, has been reviewed in several classical studies, including Cannari and D’Alessio (1992), Atkinson et al. (1995), Atkinson and Brandolini (2001), Figari et al. (2012), and references therein.

SILC survey, in the range of 23 percent for self-employment income and 44 percent for rental income. The integrated approach we develop increases the evasion rate on self-employment income to about 39 percent and that on rental income to about 65 percent. We also test for, but do not find, underreporting of employment incomes in the survey. The evasion rate on employment incomes that is estimated solely by the discrepancy method is close to 3.5 percent. The microsimulation analysis improves our ability to study the distribution profile of tax evasion. The simulations show that the evasion rates on specific income sources are generally decreasing in the various income components, which result is similar to the evidence from other studies. However, summing the component effects on total incomes increases slightly the overall evasion rates estimated for above-average income classes. This effect alters the redistributive impact of the PIT in Italy, reducing the progressivity of the tax.

The paper is organised into two main parts with several subsections. We start from the consumption-based approach to estimate income underreporting in the survey. Then we consider the microsimulation-discrepancy approach and integrate the two methods of estimating evasion. A concluding section discusses the approach further by considering also the problem of the availability of data to conduct studies that seek to analyse tax evasion.

2. The consumption-based approach to estimating income-underreporting in surveys

The consumption-based approach to estimating income-underreporting is based on the idea that differences in the income elasticity estimated in the Engel curves of goods reveal different propensities of different categories of income earners to underreport income.

Pissarides and Weber (1989; PW hereafter) and Feldman and Slemrod (2007; FS hereafter) provide two models of income-underreporting. PW developed a model of pure income based on two types of households, employees and the self-employed, in which the underreporting is related to the household's overall income. They focus on underreporting by the self-employed and assume that the degree of underreporting refers to all of their income. For their part, FS took a model of multiple income sources in which a household can underreport incomes in different amounts if they come from different sources, even if a given income source is underreported in the same proportion by different households. We take an approach closer in spirit to that of FS, even though we do not necessarily impose the latter restriction. Our approach is illustrated in the next subsection.[4]

[4] Feldman and Slemrod (2007) do not discuss the structural derivation of their model, so the exposition of our methodology follows PW more closely, with attention to the main differences from that model.

2.1. Methodology

The log-linear Engel curve for a consumption of good *C* (food in PW) is:

$$\ln C_i = \beta_0 + \beta_1 \ln Y_i^T + X_i\beta_2 + u_i, \qquad [1]$$

where i is the index for the i-th household, Y_i^T is the true household income, β_1 is the income elasticity, X_i is the matrix of vectors with household characteristics that affect the consumption decision with parameters β_2, and u_i is a white noise that may also include transitory effects of current income with respect to permanent income.[5]

A household's disposable income may come from various sources, but PW assumed a model with only two types of pure households: employees and the self-employed. Here we maintain more generality and, following FS, use a model in which the total household income is the sum of several components:

$$Y_i^T = \sum_j Y_{ji}^T. \qquad [2]$$

where Y_{ji}^T is the component from source *j* for household i.

Income-underreporting occurs when the income reported in a survey is lower than the true income. We express the relationship between true income Y_{ij}^T and reported income Y_{ij}^R as:

$$Y_{ij}^T = k_{ij} Y_{ij}^R, \qquad [3]$$

where $k_{ij} \geq 1$ is an adjustment factor that measures the extent of underreporting by household i on income *j*, with a greater k_{ij} indicating larger underreporting. In PW the adjustment factor is different from 1 only for self-employed households; it is equal to 1 for employees.

We also assume that the propensity to underreport k_{ij} is distributed as log-normal in the population and is composed of a constant term $\ln \bar{k}_j$, which depends on the type of income and by a household-specific error:

$$\ln k_{ij} = \ln \bar{k}_j + v_i, \qquad [4]$$

where $v_i \sim N(0, \sigma_v^2)$.

Alternatively, FS assume that a given income source is underreported in the same proportion by all taxpayers; that is: $\ln k_{ij} = \ln \bar{k}_j$ for all households *i*.

We maintain a specification that allows for both possibilities. Assuming that one income component, the reference income source (say $\bar{k}_1 = 1$), is correctly reported on average and that, in the more general specification, u_i and v_i are uncorrelated, equation [1] can be rewritten, substituting from equations [2] and [3], as:

$$\ln C_i = \beta_0 + \beta_1 \ln\left[Y_{1i}^R + \sum_{j \neq 1} \bar{k}_j Y_{ji}^R\right] + X_i\beta_2 + \eta_i, \qquad [5]$$

[5] Another difference between the two models is that, whereas PW maintain a specification of the Engel curve based on permanent income, FS concentrate directly on current income. The implications of the two specification strategies are discussed in more detail in section 2.4.

where $\eta_i = \beta_1 v_i + u_i$ is the new error term distributed as white noise $\eta_i \sim N(0, \sigma_\eta^2)$.

The assumption of independence between u_i and v_i is debatable, particularly when u_i is affected by transitory income components. Nevertheless, PW show in their model that the estimate of the adjustment factor is substantially unaffected, even when allowing for some correlation between the terms u_i and v_i.[6] The issue can be addressed by assuming (as in FS) $v_i = 0$ all i and/or by ignoring the difference between permanent and current income.

Equation [5] is similar to the specification studied by FS. It can be estimated by nonlinear least squares. In particular, its estimation provides directly the average adjustment factors $\bar{k}_j$ from which one obtains the average fraction of income source j that is underreported:

$$\bar{u}_j = 1 - \frac{1}{\bar{k}_j} \quad [6]$$

Moreoever, in the specification with $v_i \neq 0$, it is also possible to approximate the individual adjustment factor k_{ij} by taking u_i close to 0 so that $v_i \approx \frac{\hat{\eta}_i}{\hat{\beta}_1}$ and:

$$k_{ij} = k_j e^{\eta_i/\beta_1} \quad [7]$$

2.2. Survey data or official tax data?

Beyond the key differences concerning the definition of household income and the variables that can be underreported, the studies by PW and FS also differ in other ways. PW conducted their analysis using survey data (mainly, the 1982 wave of the Family Expenditure Survey) and found that the level of income underreporting by self-employed households in the UK averaged 33 percent (37% among blue-collar households and 29% among white-collar households). On the other hand, FS conducted their analysis using directly information from official tax returns. In particular, they used incomes from official tax registers in 1990 in the US and charitable contributions reported for tax deductions as a dependent variable. Their findings were that 35 percent of self-employment income went unreported, as did 78 percent of rents, small business income, and estate and partnership income, and 74 percent of farm income.

The use of official tax-return income as a dependent variable has the advantage that underreporting can be directly interpreted as non-compliance. However, charitable contributions reported for tax deductions as in the original approach pursued by FS or other data from the registers used

[6] The empirical specification of PW's model based on pure household income differs from equation [5] primarily in that PW estimated a model that was based on permanent income and included a dummy variable for self-employment. Assuming the same mean for the aggregate shocks to the permanent income of self-employed and employed workers, PW then showed that the coefficient of the average underreporting of self-employed households can be obtained from the coefficient on the dummy variable, with upper and lower bounds that depend on the variances and covariance of the error terms u_i and v_i. Given the assumption of multiple income sources, PW's technique based on the dummy variable and log-linear estimation cannot be applied in the present model.

as dependent variables can themselves be altered or reported untruthfully. For example, some tax-payers may exaggerate charitable contributions to benefit from tax deductions, while others may underreport them if they believe that charitable contributions that are too high—perhaps because of underreporting of income—will increase the chances of being audited. Others still may simply forget to report part or all of their donations. On the other hand, consumption data in the surveys are usually reported accurately, particularly when consumers have no motivation to misreport them. For these reasons, there is a trade-off between the use of official tax-return data and the use of survey information.

Since survey data are more accessible than official tax-return data in many countries, most studies in the field use survey data, including studies for Sweden (Schuetze 2002, Engström and Holmlund 2009, Engström and Hagen 2017), the US (Hurst et al. 2014), North Cyprus (Besim and Jenkins 2005, Elicit and Besim 2016), the UK (Cabral et al. 2014), Spain (Martinez-Lopez 2013), Estonia (Kukk and Staehr 2015, Paulus 2015a and 2015b), and Korea and Russia (Kim et al. 2017). The studies have provided consistent evidence of underreporting of self-employed households compared to employed households (in a range between 15% and 40%), although the estimates are not directly comparable because of such differences as those in dependent variables and methods of estimation.

Some studies have based their results on matching data about consumption and other information from surveys with data from official tax-return incomes. Matching survey data with official tax-return data is appealing because it offers the possibility of interpreting income underreporting as tax non-compliance while still using measures from a consumption survey as dependent variables, rather than expenditures reported for tax deductions, which are prone to misreporting. These studies include an analysis for Finland (Johansson 2005), two investigations for Sweden (Engström and Holmlund 2009, Engström and Hagen 2017), and two for Estonia (Paulus 2015a and 2015b). Paulus (2015b) compares the underreporting estimated from survey-reported income data with official tax-return data. The analysis confirms that the extent of underreporting is much higher for official tax-return incomes than it is for survey-reported incomes (an average of 20% underreporting for the self-employed using survey data versus 56% underreporting for the self-employed using official tax-return data). Unfortunately, these types of data are not available for the present study on Italy.

2.3. Data, income categories, and the dependent variable

Our analysis investigates underreporting of income using the IT- SILC, which is also used by the microsimulation model BETAMOD in the second part of the paper.

We use the IT-SILC 2011 cross-sectional wave, which has a considerably larger sample size than the rotating longitudinal component does. The interviews are structured into a household questionnaire and an individual questionnaire administered to all household members age sixteen and older. The household part collects information on the households' composition, accommodation, housing costs, and economic circumstances, such as household savings, debt, means-tested benefits, and children's income, while the individual part covers information on individual incomes differentiated by source, and other information, including education, health and occupation.

The self-reported information on income sources refers to employment and self-employment incomes, pensions, unemployment and disability or incapacity benefits, rental income from immovable properties, partnerships, financial investments, and other capital income.[7]

The household part of the IT-SILC does not include household expenditures on food, which has been used as a dependent variable in most studies on income-underreporting, but it includes a rich battery of expenditures for running a household that have been tested as a way to estimate underreporting (Cabral et al. 2014, Hurst et al. 2014, Paulus 2015b). The dependent variable is based on an aggregate of home-related expenditures, including costs for heating, electricity, gas and other fuels, water, and condominium fees.

We take as our main reference income category the aggregate of pensions and unemployment benefits and other state benefits that are subject to withholding and can be easily identified and recalled. We estimate potential underreporting for two main income categories: income from self-employment and rental income from immovable properties, partnerships, and other capital income, to which we refer as "rents" henceforth.

The use of pension income and other state benefits as a reference category is not standard in the literature. In particular, most previous analyses that were based on survey data and that followed the approach of pure income households have taken employee households as the reference income households. A problem with this choice is that employees may themselves underreport income. We seek to verify this possibility in the data set.[8]

Moreover, since the consumption-based approach rests on the hypothesis that the propensity to consume does not vary based on the income source, one may also wonder whether households with only pension income spend their incomes on home utilities as other income earners. We conduct some sensitivity checks to investigate this assumption as well. We nevertheless emphasize again that our approach, rather than testing for underreporting of specific income categories, is driven by integrating the consumption-based approach with the microsimulation discrepancy approach to produce an overall estimate of evasion across the whole population.

For this reason, unlike most studies which have restricted their samples to homogeneous households, our investigation is based on a full data set using the 2011 cross-sectional wave of IT-

[7] More recently, EUROSTAT and the national statistical institutes that manage EU-SILC, including ISTAT for IT-SILC, have broadened the use of administrative data to check and control the income data that is collected through surveys (Consolini and Donatiello 2015). This type of control will improve the quality of data for research purposes, particularly if both types of income information—that from tax registers and that from surveys—are reported in the EU-SILC. (We discuss this point more in the conclusion.)

[8] This possibility is usually associated with private employment income since civil servants have virtually no opportunity to underreport the income they receive from the public sector (e.g. Besim and Jenkins 2005, Ekici and Besim 2016, Paulus 2015a). However, the IT-SILC does not indicate whether all of the employment income reported in the survey by public employees comes from the work in the public sector or some comes from second/part-time jobs in the private sector. Therefore, it is not possible to use the distinction between public employment income and private employment income in the present approach.

SILC, which includes 19,043 households with positive home-utility expenditures.

2.4. Estimation and results

We estimate several specifications of equation [5], taking expenditures for home utilities as the dependent variable and using the nonlinear least-square estimation.

The literature has considered various income measures as independent variables in estimating underreporting. FS estimated underreporting using current income sources,[9] while PW estimated underreporting using instrumental variables to reduce the effect of measurement errors and transitory components of income. As an alternative method, some studies have proposed constructing measures of permanent income and have used them directly as dependent variables. Such approaches have typically been based on the average over some years of income obtained from survey data with a panel dimension (as in, e.g., Besim and Jenkins 2005, Hurst et al. 2014, Kim et al. 2017).[10]

We estimated our main model using instrumental variables (IV) that correspond to model (1) in Table 1. Suitable instruments respecting identifying restrictions have been used for employment incomes and self-employment incomes. In particular, the main instruments include variables for individual characteristics and human capital expressed as dummies for education, occupation and economic sector, physical assets, and geographic region. Various other covariates, including additional households characteristics, homes characteristics, and dummies for geographic area, are included in the estimation of equation [5].

The IV estimates of model (1) are consistent with the hypothesis of underreporting. In particular, the adjustment factors on self-employment income and rent income are, respectively, $\hat{\hat{k}}_{self} = 1.298$ and $\hat{\hat{k}}_{rent} = 1.773$. Both coefficients are significantly greater than 1 at the 5% level (one-tail test). The corresponding underreporting rates are $\hat{\hat{u}}_{self} = 22\%$ for self-employment and $\hat{\hat{u}}_{rent} = 44\%$ for rents.

In model (2) of Table 2 the same specification is estimated using current incomes. The estimates of the underreporting coefficients are a bit lower than those with IV and are not significantly different from those in model (1). The result is consistent with the notion that transitory income fluctuations may bias the estimates' precision. Nevertheless, the IV and the current income estimates of income elasticity β_1 are similar (0.086 and 0.093, respectively). This result imbues IV estimation with some confidence since an occasional criticism of IV in this literature is that it may overestimate the income elasticity β_1.

[9] Nevertheless, FS use several controls to check for spurious regressions and biased estimated coefficients that are due to measurements errors and find no evidence of such issues in their estimation of the effects.

[10] Another approach was used in Kukk and Stare (2015), who used a survey that includes questions on current income and regular income and took the latter as a measure of reported permanent income. A summary of the discussion is in Paulus (2015b).

Model (3) in Table 1 adds the possibility of underreporting on employment income. The point estimate on employment income is $\hat{\bar{k}}_{empl} = 1.031$, which is not significantly different from 1 at statistical level. Thus, the model rejects the hypothesis that employment income is underreported.

In model (4) of Table 1 we test for the sensitivity of excluding households with only pension incomes from the regression. The results appear to be robust to the exclusion.

Overall, the results speak in favour of accepting model 1 as the specification for estimating income underreporting in the IT-SILC database. We also recall that the specification of the model consents to two interpretations for the estimates of $\bar{k}_{self}$ and $\bar{k}_{rent}$. Under the first interpretation, which is more consistent with FS, the two adjustment factors are assumed to be the same for all taxpayers. Alternatively, $\bar{k}_{self}$ and $\bar{k}_{rents}$ can be considered the mean values of the individual adjustments factors that are determined according to equation [4]: $\ln k_{ij} = \ln \bar{k}_j + v_i$, for income source j and idiosyncratic error v_i. Unfortunately, we cannot identify the two error terms v_i and u_i that enter the regression, which would have allowed us to obtain a precise estimate of the k_{ij}s . However, assuming u_i close to 0 allows us to approximate the k_{ij}s from equation [7] as $\hat{k}_{ij} = \hat{\bar{k}}_j e^{\hat{\eta}_i/\hat{\beta}_1}$.

The distributions of the adjustment factors $\hat{k}_{ij}$ and of the corresponding underreporting rates $\hat{u}_{ij}$ are shown separately for self-employment incomes and rents in Fig. 1.[11] About 53 percent of taxpayers with self-employment income underreport. The modal underreporting rate among those who underreport is 26% while a few taxpayers (almost 7%) underreport more than 50 percent of this type of income. Underreporting is higher for rental income, with 56 percent of taxpayers earning this type of income who underreport it. The modal underreporting rate among those who underreport is 41 percent and about 22 percent taxpayers with this type of income underreport more than half of it.

3. Evasion rates and microsimulation analysis with correction for underreporting

Income-underreporting in a survey may include only a part of the non-compliance committed by taxpayers. We seek to integrate income-underreporting in the discrepancy approach in order to estimate the total tax evasion.

The discrepancy approach computes evasion rates by comparing the income distribution from surveys with the distribution based on the official tax-return data. However, surveys and official tax-return data do not contain the same information, so the method requires adjustments in order to

[11] The plotted values are effectively based on the transformation $\hat{k}_{ij} = \left[\max\left(\hat{\bar{k}}_j e^{\hat{\eta}_i/\hat{\beta}_1}, 1\right)\right]^{\hat{r}_j}$, where $\hat{r}_j$ are exponents for the normalization such that the resulting distributions have the estimated mean $\hat{\bar{k}}_j$.

ensure that the income variables and the population that underlies the income distributions from the surveys and from the official data are consistently defined and comparable. In the surveys the respondents are typically asked for their disposable income, whereas official tax-return data usually comes in the form of tables of gross income and taxes.[12] Surveys and official data also contain different types of measurement errors; for example, measurement errors in surveys may be due to factors like imprecise answers, lapses of concentration, or inexact recording by the interviewers. These types of errors are less likely to occur in official tax-return data (Paulus 2015a and references therein). A related issue is that official tax-return data refer to the population of individual taxpayers, whereas most surveys use a sample of representative households that may contain sample errors and be affected by non-response rates. (See the discussion and references in Fiorio and D'Amuri 2006 and Marino and Zizza 2008.)

The literature that has applied the discrepancy method has followed various procedures to address these issues.[13] Our application of the discrepancy approach is based on microsimulation analysis[14] and is part of a microsimulation model called BETAMOD.

3.1. Evasion rates in BETAMOD

BETAMOD is a model for the Italian PIT (Irpef, with regional and municipal surtaxes), which includes in the simulation an internal procedure to check for consistency between simulation results that are based on survey data and those that are based on official tax-return data. It is through this procedure that the discrepancy approach estimates evasion rates.

The model works through various steps and modules in an iterated process.[15] The simulations begin with individual disposable incomes and other households' information from IT-SILC and other surveys and adjusts tax expenditures based on family and personal characteristics. First-round simulations are then conducted to convert net income to gross income. Unlike models that impute tax expenditures though calibration with aggregate fiscal data by income class, calibration weights are estimated in order to match population totals with administrative counts of taxpayers. The model then produces output that includes estimates of individual evasion rates by comparing the gross

[12] More recently, EU-SILC has started to include a variable for household gross income obtained with a multi-country microsimulation model devised by the University of Siena to perform the gross-net conversion of the incomes (Betti, Donatiello and Verma 2011).

[13] In most studies the comparison between survey data and official tax-return data has been made in terms of after-tax incomes, with computations in the official tax-return data based on imputations of tax and benefits outside of the microsimulation model (Bernasconi and Marenzi 1997, Mantovani and Nienadowska 2007, Fiorio and D'Amuri 2006, Matsaganis et al. 2010, Marino and Zizza 2008). However, this method has shortcomings, including its basis on unchecked imputations and tendency to overestimate the evasion rates since it computes them as the ratio between evaded income and net income instead of true gross income.

[14] The use of micro-simulation models in economics for public decision-making has developed enormously in the last thirty years and is now a widely employed method of analysis that uses various techniques with theoretical backgrounds. (See discussions and references in Mitton et al. 2000, Bourguignon and Spadaro 2006, Immervoll et al. 2007, Figari et al. 2015.)

[15] A full description of the model is in Albarea et al. (2015).

income obtained with disaggregated data from official tax-return data. Round-specific convergence measures are also assessed in term of consistency with official tax-return data, and the model is iterated until convergence is achieved; that is, the iterations stop when the reported levels of income estimated by the model using estimated taxes, tax allowances, and calibration weights do not differ significantly from the official tax-return data at both the aggregate level and the subgroup level, with the latter defined by main source of income and geographic area.

The evasion rates are computed as follows. Aggregate evasion rates are computed to compare simulated true gross incomes with administrative tax data on reported income. As administrative data are provided in the aggregate by main income source (employment income, pensions, self-employment income, rental income from immovable property) and separately by geographical area (northwest, northeast, central, south, and isles), the model applies the RAS technique to obtain the joint distribution of reported income in both dimensions,[16] resulting in a 4×4 matrix of average evasion rates by income type and geographical area. A distribution income profile of tax evasion is estimated for each area-by-income-type stratum. Then we refine the stratification by expanding the 4×4 strata to account for the profile of tax evasion by income class. In more detail, each area-by-income-type stratum is expanded into thirteen classes of true gross income to obtain sixteen income profiles of tax evasion.[17] The design of each evasion-by-income profile results from an optimizing procedure that minimizes the distance between simulated and official tax-return income. The result is a (4x4)×13-dimension matrix of evasion rates by main income source, geographical area, and true gross income level.

Even if the evasion rates are listed by matrix cells, individual evasion rates are determined for each micro-unit, depending on the composition of individual income in terms of income source. These types of composition-effects are important for the overall distribution impact of tax evasion, and ignoring them may lead to a substantial underestimation of the regressive impact of tax evasion. (See additional discussion in Mantovani and Nienadowska 2007, and Matsaganis et al. 2010.) Moreover, with BETAMOD it is possible to consider procedures that would obtain even more personalised evasion rates. One possible procedure can result from integrating the consumption based-approach into BETAMOD.

3.2. Integrating the consumption-based approach into BETAMOD

BETAMOD runs on the same cross-sectional component of IT-SILC 2011 that is used in the consumption-based approach. Since IRPEF is reported on an individual basis, BETAMOD uses the household information from IT-SILC 2011 to conduct simulations at the individual level. The data set is also

[16] The RAS algorithm is an iterative proportional fitting procedure proposed by Bacharach (1965) that estimates the joint distribution of two or more variables from their marginal distributions.

[17] The gross income classes (in thousands of euros) are: 0-5, 5-7.5, 7.5-10, 10-12, 12-15, 15-20, 20-26, 26-29, 29-35, 35-40, 40-50, 40-50, 50-75, and >75.

enriched with additional information that is not available in IT-SILC but that comes from two other surveys: the 2010 Survey on Households Income and Wealth (SHIW) released by the Bank of Italy (2012), which is used to compute cadastral values and tax relief for imputation of payments of insurance premiums and other home-refurbishments expenditures, and the 2013 MULTISCOPO Survey on Health Conditions and the Use of Health Services (ISTAT 2014), which is used to compute tax relief for healthcare expenditures. Imputations are performed using statistical matching techniques, where SHIW and MULTISCOPO individuals have provided the information that is missing from IT-SILC. (See additional details in Albarea et al. 2015.)

The other main modifications conducted here concern corrections for income underreporting in IT-SILC. In particular, the evidence obtained from the consumption-based method rejects the assumption that income is truthfully reported in the survey data. Accordingly, incomes of the micro-units from IT-SILC that represent the input of BETAMOD have been corrected using the underreporting rates estimated by the consumption-based approach. In particular, we conducted three simulations: the first simulation (simulation A) was run as the benchmark and was conducted with the original incomes of the micro-units from IT-SILC. The other two simulations applied two procedures to correct the input data. Simulation B multiplied the various income components j of individual i by the mean adjustment factor $\hat{\hat{k}}_j$, so the income of micro-unit i used as input by BETAMOD is:

$$Y_i^B = \sum_j \hat{\hat{k}}_j Y_{ij},$$

where, as we recall, $\hat{\hat{k}}_{self} = 1.298$ and $\hat{\hat{k}}_{rent} = 1.773$. Simulation C used a similar imputation but used the personalized adjustment factors $\hat{k}_{ij}$ obtained at the end of Section 2.4 (in Fig. 1):

$$Y_i^C = \sum_j \hat{\hat{k}}_{ij} Y_{ij}$$

Therefore, whereas in simulation B all micro-units were assumed to have the same propensity to underreport the income of a given source, simulation C attached different underreporting rates to different individuals and distributes the total evasion that was estimated with the discrepancy approach. Mainly, the total tax-evasion of each cell of the matrix of the evasion rates is distributed among the individuals of the same cell using the individual adjustment factors $\hat{\hat{k}}_{ij}$. The hypothesis here is that individuals' propensity to underreport in the survey is correlated with their propensity to underreport income in their official tax-return reports to the tax authorities, even if the scale of underreporting in the surveys and that of cheating to the tax authorities differ. In addition, since (as will be shown) the discrepancy approach also provides evidence of evasion on employment income, the same procedure is used to distribute evasion by employees.[18]

[18] Since the adjustment factor for employment income is $k_{emp} = 1$, the individual adjustment factors used to distribute evasion are $\hat{\hat{k}}_{i,empl} = \left(1 \cdot e^{\hat{\eta}_i/\hat{\beta}_1}\right)^{\hat{r}_{empl}}$.

3.3. Results of the simulation and aggregate measures of tax evasion

In presenting the results of the simulations, we first give evidence on the ability of BETAMOD to simulate the PIT and local income taxes in Italy for 2010 and provide evidence on aggregate evasion measures. Then we show the distribution profiles of the estimated evasion rates.

A major validation criterion for simulations is consistency with external data sources (Atkinson et al. 1988, D'Amuri and Fiorio 2006, Figari et al. 2012). Fig. 2 compares the distribution of taxpayers (Fig. 2a) with that of reported income (Fig. 2b) by classes of reported income between the official tax-return data and the simulations. The similarity between the distributions simulated by BETAMOD and the official tax-return distribution is striking. In fact, even for the two central classes of reported incomes, which comprise the largest fractions of taxpayers and the greatest amount of reported income (15-20 and 20-26 thousand euros), the difference between the percentages from the official tax-return data and the simulations is around 1 percent. The consistency between the simulated and the official distributions is important as, among other things, it verifies the fitness of BETAMOD to conduct distribution analyses.

Table 2 provides aggregate quantifications for the main components of the Irpef and local taxes and compares them with official data. The results include the re-weighting of the IT-SILC data by BETAMOD. All simulations provide continued evidence of the consistency of the simulations with the official tax-return data: the percentage differences between the simulated results for the main Irpef and local tax components with the official tax-return data are between -0.5 percent and 3 percent in most cases. The only difference that is more than 5 percent arises in all simulations with respect to the number of individuals with positive gross tax liability, which is likely to depend on the model's imputation of tax deductions and result in a larger number of individuals with positive taxable income in BETAMOD.

Table 2 also reports the total gross incomes obtained by the three simulations to measure tax evasion. Simulation A, conducted without correcting the IT-SILC data for underreporting, estimates that on the aggregate slightly more than €61.4 billion in gross income escapes the tax authorities, corresponding to an evasion rate of 7.25 percent. Correcting for underreporting in simulations B and C raises the estimates to €124.5 and €132.1 billion in total gross income evaded, respectively, corresponding to evasion rates of 13.69 percent and 14.27 percent, respectively in the two simulations.

These numbers attest to the relevance of the problem of underreporting in the survey to the ability to quantify the dimension of evasion. In particular, an estimated evasion rate of around 7 per cent is low compared to that obtained by other studies and methodologies. (See references and discussion in Marino and Zizza 2008.) The rates estimated by simulations C and B are more in line with the literature. For example, studies reviewed in Giovannini (2001) estimated tax evasion by means of unreported income in the range of 13-25 percent of GDP.[19]

[19] Comparisons with other studies are not always easy given the different periods of estimation, focuses of the measures, and ways of conducting comparisons. Studies that have applied the discrepancy approach to micro data have

The small differences between simulations B and C in the estimates of total evasion occur because the individual and total true gross incomes are obtained by means of the net-to-gross procedure, which is affected by the initial income distribution. The average and personalized procedure used to correct for underreporting has a minor impact at the aggregate level, confirming the neutrality of the procedure in terms of affecting the total gross income and total evaded income estimated by BETAMOD (less than 1% on the overall evasion rate).

Table 3 reports the average evasion rates by income source and geographic area and documents the impact of correcting for underreporting in the estimation of evasion rates. The estimates from all simulations confirm that tax evasion on employment income (around 3.3-3.8% in all simulations) is lower than that on self-employment and rental income from immovable property. In fact, the estimation of evasion rates on these two income sources, already substantial in simulation A (23.6% on self-employment and 29.0% on rental income), increase in the simulations that correct for underreporting, to 37.4% and 39.4% for self-employment in simulations B and C, respectively; and to 62.2% and 65.9% for rental income in simulations B and C, respectively. The percentages in Table 3 also reveal differences among geographic areas: in particular, all simulations identify the south of Italy as having systematically higher evasion rates, followed by the northeast. The correction for underreporting does not modify appreciably the proportional differences among the geographic areas, supporting the hypothesis that people’s propensity to underreport incomes in surveys is consistent (even if on a lower scale) with their inclination to conceal income from the tax authorities.

Table 4 shows the losses in tax revenues (tax gap) that are due to tax evasion. They are obtained by simulating scenarios in which taxpayers fully report their incomes. The total tax gap is about €16.5 billion in simulation A, €37.5 billion in simulation B, and €38.6 billion in simulation C. In all simulations the larger part of the tax gap is caused by evasion on self-employment income, which in the simulations B and C which correct for underreporting is close to €21 billion. In these latter simulations the tax gap on employment income is higher in simulation B and a bit lower in simulation C because of the differences in the tax evasion profiles estimated by BETAMOD in the two simulations. On the other hand, the tax gap on rental income is higher in simulation C than it is in B (€14.7 and € 12.6 billion, respectively) and much lower in simulation A, which does not correct for underreporting (about €3.3 billion).

Finally, the tax gaps estimated on self-employment in simulations B and C are consistent with the tax gap that a recent official report (*Ministero dell'Economia e delle Finanze* - MEF 2016) estimated

often reported lower evasion rates than have studies that have applied the approach at the macro level. (See references and discussion in Giovannini, 2011.) For example, Bernasconi and Marenzi (1997) compared the micro data recorded in the Survey of Household Income and Wealth (SHIW) with the official tax-return data for the fiscal year 1991 and estimated an overall evasion rate of 15 percent, whereas Bernardi and Bernasconi (1996) used the macroeconomic approach and estimated a total evasion rate of 26 for the same year. The evasion rate of 7.25 percent that was estimated in simulation A seems particularly low, even with respect to other studies that have applied the discrepancy approach without correcting for underreporting (e.g. Bernasconi and Marenzi 2001, or the analyses from Fiorio and D'Amuri 2006 and Marino and Zizza 2008, both of which estimated an overall evasion rate between 13.5 and 15 percent). These studies are however not fully comparable with the present analysis since all of the earlier investigations that used the discrepancy approach on micro data measured evasion in terms of after-tax incomes.

for the same fiscal year and on the same income source. In particular, the official tax gap on self-employment incomes estimated by the report using a macro approach is €20.1 billion.[20] The similarity between the estimates here and those from MEF is another indication of the importance of correcting for income-underreporting in the discrepancy approach at the micro level.

3.4. Distribution effects of tax evasion

An advantage of microsimulation analysis is that it permits one to study the distributive effects of evasion. Fig. 3 reports the evasion rates by income source and gross income class obtained with the three simulations. A feature common to the three simulations is that the evasion rates computed for employment incomes, self-employment incomes and rental incomes all have a negative gradient. This evidence is consistent with studies that have shown evasion rates generally decreasing in the three components of income (Bernasconi and Marenzi 1997, Fiorio and D'Amuri 2006). However, unlike these previous studies, the present analysis finds in all simulations comparatively flatter gradients of the evasion rate on employment income, possibly because the evasion rates computed by BETAMOD are over gross income, while in earlier works' estimates are usually based on net income. Moreover, the evasion rate for total income remains decreasing overall only in simulation A (similar to previous works), while in simulation B and especially in simulation C the evasion rate for total income is first decreasing (reaching the lowest value around 10% at a gross income level of about 22 thousand euros) before slightly increasing to become almost constant (around an evasion rate of about 20%) on classes with very high gross income.

These impacts on total income are due to the composition effects illustrated in Fig. 4. The figure shows the total amount of unreported income by income class and income source in the three simulations. In simulation A, despite the decreasing profile of the evasion rates, most evaded income is from taxpayers in the central-income classes (between 12,000 and 35,000 euros) whose gross income is from self-employment and employment. On the other hand, in simulations B and C the highest amount of evaded income comes from income earners in the highest-income classes whose gross income is mainly from self-employment and rentals.

By reducing reported income, tax evasion causes a modification in the actual tax schedule with respect to theoretical ones (Fig. 5), modifying the redistributive effect of the tax schedule, which can change its progressivity impact and have horizontal inequity effects and reranking effects. Table 5 reports a set of standard inequality indices to evaluate the redistributive impact of tax evasion in simulations A and C. (Simulation B is midway between simulations A and C, so it is omitted.)[21] In

20 The report, "Relazione sull'economia non osservata e sull'evasione fiscale e contributiva" (MEF 2016), did not produce assessments for the tax gap on rental income, while it estimated a tax gap on employed workers of about €3.9 billion that arose from irregular jobs. Unfortunately, the official document did not report estimates of tax evasion that was associated with the various tax gaps.

21 We report here the indices for the distributions of individual income; alternative measures with similar results can be obtained by computing the indices for the distributions of households with equivalent income.

simulation A, tax evasion makes the distribution of reported income to appear more unequal than it is: the Gini index of the distribution of reported income with evasion is higher than the index of the distribution without evasion (0.432 versus 0.414). On the other hand, the difference in inequality between the distributions of net incomes with and without evasion is negligible (0.363 versus 0.364).[22] Regarding simulation C, the distributions of true gross income and of reported incomes both with and without evasion (Gini 0.435 and 0.453, respectively) are more unequal than the corresponding distributions in simulation A. This shows that ignoring underreporting tends to underestimate inequality in the pre-tax income distributions. Moreover, simulation C shows that evasion also increases inequality in the distribution of after-tax incomes. The concentration index for the distribution with evasion (0.394) is higher than the index for the theoretical distribution without evasion (0.380); the Reynolds-Smolensky index is lower in the tax simulation with evasion, mainly because of a strong reduction in the average tax rate (by 4 percentage points). Moreover, evasion also causes a positive reranking effect.

Finally, we consider the distributions of the evasion rate across income sources for the three simulations (Fig. 6). The results of simulation A, conducted without correcting for underrreporting, have two unrealistic features: that for all types of income and for total income the evasion rates are very low compared to other studies; and that all taxpayers with either employment income, self-employment income or rental income evade their incomes to some extent at least, so the fraction of fully compliant taxpayers (22.7%) is only those whose total income comes from pensions.

Simulation B partially corrects for the first feature, but it produces some other unrealistic effects and leaves the second feature unaffected. In particular, the evasion rates estimated on self-employment and rental incomes are higher in simulation B than they are in simulation A, producing average evasion rates that are consistent with common perceptions and evidence from other sources. However, the way in which the input data are corrected for underreporting in simulation B (i.e. by multiplying the income component j by the mean underreporting factor $\bar{k}_j$), along with the fact that the imputation of the evasion rates to the micro units continues to be given equally to all individuals in the matrix cells, leaves all self-employed workers, rental-income earners, and employed workers as tax-evaders and also the distribution profiles of the evasion rates concentrated into a few evasion rates (with more than 66% of self-employed income earners having evasion rates higher than 50%, and more than 83% of those with rental income having evasion rates greater than 75%).

Simulation C produces more realistic profiles of the distributions, with more substantial fractions of taxpayers who fully report, particularly among employed workers (47.4%). The distributions of the evasion rates for self-employed earners and rental-income earners also become less concentrated than they are in simulation B, especially in the case of the self-employed (with almost 12% of them in full compliance, about 40% with evasion rates lower than 50%, and about 48% with evasion

[22] By comparing the differences in the distributions of gross incomes, one could then point out that, according to simulation A, the impact of evasion is a moderate increase in the redistributive effect of taxation, as is also at least partially indicated by the Kakwani index for the schedule with evasion, which is higher than that without evasion, even if the impact of this difference for the Reynolds-Smolensky index is completely counterbalanced by the reduction in the average tax rate with evasion.

rates higher than 50%). Composing the effects on total income produces a distribution of evasion rates that also differs from the previous simulations, with the greatest fraction of taxpayers who fully comply (43.2%), a substantial fraction (32.3%) who fail to report less than 20 percent of their income, and smaller and decreasing proportions of taxpayers evading at higher rates (3.3 percent of taxpayers with evasion rates on total income that are higher than 80%).

4. Conclusion

Measuring the amount of tax evasion and its impact on the economy is a complex process, so several methods are required to triangulate the size and the effects of the black economy (Cabral et al. 2014).

In this paper we have proposed an approach that integrates the consumption-based method of estimating income-underreporting in surveys with a microsimulation-based discrepancy analysis to determine evasion rates by comparing survey data with official tax-return data. We have used the consumption-based method to estimate income-underreporting of self-employment incomes and rental incomes from capital and immovable properties in the IT-SILC, the Italian part of the EU-SILC database. We have also found some signs of underreporting of employment income, but not at a statistically significant average level. Using the discrepancy method, we have found that there is a substantial amount of tax non-compliance that occurs in addition to income underreporting, and we have used micro data corrected for underreporting to estimate the distribution profile of tax evasion, distinguishing between simulations based on average profiles and those based on individual profiles.

The quality of the data used in the empirical investigations on tax evasion is also important. To conduct our analysis we have used the 2011 wave of IT-SILC, where income information is still based primarily on survey information. There is now a growing discussion by Eurostat and national statistical offices to consider the use of official tax-return data for income in the context of the EU-SILC (Eurostat 2013). There are benefits and costs for using official tax-return data, and the extent to which the data are actually used in practice varies widely across countries. The likely best practice is to use official tax-return data to complement data collected through surveys, rather than as a substitute.

Datasets that combine survey data with official tax-return records could greatly benefit the study of tax evasion, particularly when income information from the two sources are linked at individual level.[23] In that case, the two income sources could be used to improve the identification of

[23] This type of dataset is rare, often because of privacy concerns, but there are no technical reasons for inhibiting methods that allow exact matching of individual records and at the same time preventing identification of the individual (Boeri and Pellizzari 2007, Trivellato 2007).

underreporting for the purpose of tax evasion—including in analyses of income-underreporting among employees (e.g. as in Paulus 2015a)—and separate it from other types of measurement errors.

These benefits would come primarily from the consumption-based method and would help to align the estimates of income underreporting to the total evasion rates that are computed with respect to disposable income.

The microsimulation discrepancy method and the integrated approach proposed here would nevertheless remain valid: for deriving gross incomes and double-checking with the discrepancy approach for any tax evasion remaining even after conducting the grossing-up procedure based on a detailed microsimulation system; for computing the tax gap that is due to tax evasion with the same accurate microsimulation model used for tax and benefits; and for analysing the distribution impact of tax evasion, along with the possibility of informing the microsimulation analysis of individual propensities to evade taxes that are estimated by the consumption-based method.

References

Albarea, A., Bernasconi, M., Di Novi, C., Marenzi, A., Rizzi, D., and Zantomio, F. (2015). Accounting for tax evasion profiles and tax expenditures in microsimulation modelling. The BETAMOD model for personal income taxes in Italy. *International Journal of Microsimulation*, 8, 99-136.

Alm, J., and Embaye, A. (2013). Using dynamic panel methods to estimate shadow economies around the world, 1984–2006. *Public Finance Review*, 41(5), 510-543.

Atkinson, A. B., and Brandolini, A. (2001). Promise and pitfalls in the use of ''secondary'' data-sets: Income inequality in OECD countries as a case study. *Journal of Economic Literature*, 39(3), 771–799.

Baldini, M., Bosi, P., and Lalla, M. (2009). Tax evasion and misreporting in income tax returns and household income surveys. Politica Economica, XXV (3), 333– 348.

Bernardi, L. e Bernasconi M. (1997). L'evasione fiscale in Italia: evidenze empiriche. *Il fisco*, n. 38, pp. 19-36.

Bernasconi, M. and Marenzi, A. (1999). Gli effetti redistributivi dell'evasione fiscale in Italia. *Ricerche quantitative per la politica economica 1997*. Convegno CIDE-SADIBA, Banca d'Italia, novembre.

Betti, G., Donatiello, G., and Verma, V. (2011). The Siena microsimulation model (sm2) for net-gross conversion of EU-SILC income variables. *International Journal of Microsimulation*, *4*(1), 35-53.

Besim, M., and Jenkins, G. P. (2005). Tax compliance: when do employees behave like the self-employed? *Applied Economics*, 37 (10), 1201–1208.

Boeri, T., and Pellizzari, M. (2007). La deontologia di chi produce e detiene dati statistici: dalla possibilità alla certezza dell'accesso. *Statistica*, *63*(4), 649-662.

Bourguignon, F., and Spadaro, A. (2006). Microsimulation as a tool for evaluating redistribution policies. *Journal of economic Inequality*, *4*(1), 77-106.

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

Cannari, L., Ceriani, V., and D’Alessio, G. (1997). Il recupero degli imponibili sottratti a tassazione. *Ricerche quantitative per la politica economica 1995*. Convegno CIDE-SADIBA, Banca d’Italia, novembre.

Cabral, A.C.G., Kotsogiannis, C., and Myles, G. (2014). Self-Employment Underreporting in Great Britain: Who and How Much? Tax Administration Research Center Discussion paper 010-14.

Cozzolino, M., and Marco, M. D. (2015). Micromodelling Italian Taxes and Social Policies. *Rivista di statistica ufficiale*, *17*(2), 17-26.

Consolini, P., and Donatiello, G. (2015). Multi-source data collection strategy and microsimulation techniques for the Italian EU-SILC. *Rivista di statistica ufficiale*, *17*(2), 77-96.

Ekici, T., and Besim, M. (2016). A measure of the shadow economy in a small economy: evidence from household-level expenditure patterns. *Review of Income and Wealth*, *62*(1), 145-160.

Engström, P., and Holmlund, B. (2009). Tax evasion and self-employment in a high-tax country: evidence from Sweden. *Applied Economics*, 41(19), 2419–2430.

Engström, P. and Hagen, J. (2017). Income underreporting among the self-employed: a permanent income approach. *European Economic Review*, 92, 92-109.

EUROSTAT (2013). *The Use of Registers in the Context of EU-SILC: Challenges and Opportunities,* Jäntti, M., Törmälehto, V. M., and Marlier, E. (eds.). Eurostat *2013 Edition*. Publications Office.

Feldman, N., and Slemrod, J. (2007). Estimating tax noncompliance with evidence from unaudited tax returns. *The Economic Journal*, 117, 327–352.

Figari, F., Iacovou, M., Skew, A. J., and Sutherland, H. (2012). Approximations to the truth: comparing survey and microsimulation approaches to measuring income for social indicators. *Social Indicators Research*, 105(3), 387-407.

Figari, F., and Paulus, A. (2015). The distributional effects of taxes and transfers under alternative income concepts: the importance of three “I”. *Public Finance Review*, 43(3), 347-372.

Figari, F., Paulus, A., and Sutherland, H. (2015). *Microsimulation and Policy Analysis*, in “Handbook of Income Distribution”, Volume 2B, edited by A. B. Atkinson and F. Bourguignon, Elsevier, 2141-2210.

Fiorio, C. V., and D’Amuri, F. (2006). Tax evasion in Italy: An analysis using a tax-benefit microsimulation model. *The IUP Journal of Public Finance*, 4, 19–37.

Giovannini, E. (2011). Economia non osservata e flussi finanziari. Rapporto finale. Ministero dell’Economia e delle finanze, Roma.

Hurst, E., Li, G., and Pugsley, B. (2014). Are household surveys like tax forms: evidence from income underreporting of the self-employed. *Review of Economics & Statistics*, 96(1), 19–33.

Immervoll, H., Kleven, H. J., Kreiner, C. T., & Saez, E. (2007). Welfare reform in European countries: a microsimulation analysis. *The Economic Journal*, *117*, 1-44.

Johansson, E. (2005). An estimate of self-employment income underreporting in Finland. *Nordic Journal of Political Economy*, 31, 99–109.

Kim, B., Gibson, J., and Chung C. (2017). Using panel data to estimate income under-reporting by the self-employed. *The Manchester School*, 85, p. 41-64

Kukk, M., and Staehr K. (2015). Identification of Households Prone to Income Underreporting Employment Status or Reported Business Income? *Public Finance Review*, 1-29

Lyssiotou, P., Pashardes, P., and Stengos, T. (2004). Estimates of the black economy based on consumer demand approaches. *The Economic Journal*, 114 (497), 622–640.

Mantovani, D., and Nienadowska S. (2007). The distributive impact of tax evasion in Italy. Materiali di discussione no. 575, Dipartimento di Economia Politica, Università degli Studi di Modena e Reggio Emilia.

Marenzi, A., (1996). Prime analisi sulla distribuzione dell'evasione IRPEF per categorie di contribuenti e per livelli di reddito, in Rossi, N. (ed.), *Competizione e Giustizia sociale, III rapporto CNEL sulla distribuzione e redistribuzione del reddito in Italia*, Il Mulino, Bologna.

Marino, M. R., and Zizza R. (2008). L'evasione dell'Irpef: una stima per tipologia di contribuente. *Banca d'Italia, Servizio Studi di Struttura economica e finanziaria*.

Martinez-Lopez, D., (2013). The underreporting of income by self-employed workers in Spain. *SERIEs*, 4 (4), 353–371.

Matsaganis, M., Benedek, D., Flevotomou, M. , Lelkes, O., Mantovani, D., and Nienadowska S. (2010). Distributional implications of income tax evasion in Greece, Hungary and Italy. MPRA, Munich Personal RePEc Archive.

MEF (2016). Relazione sull'economia non osservata e sull'evasione fiscale e contributiva. Ministero dell'Economia e delle Finanze, Roma.

Mitton, L., Sutherland, H., & Weeks, M. (2000). *Microsimulation modelling for policy analysis: challenges and innovations*. Cambridge University Press.

Paulus, A. (2015a). Tax evasion and measurement error: An econometric analysis of survey data linked with tax records. ISER Working Paper Series, no. 2015-10, Institute for Social and Economic Research, University of Essex, Colchester, UK.

Paulus, A. (2015b). Income Underreporting Based on Income-expenditure Gaps: Survey vs Tax Records. ISER Working Paper Series, no. 2015-15, Institute for Social and Economic Research, University of Essex, Colchester, UK.

Pissarides, C.A., and Weber, G. (1989). An expenditure-based estimate of Britain's black economy. *Journal of Public Economics*, 39, 17–32.

Schneider, F. (2005). Shadow Economies Around the World: What Do We Really Know? *European Journal of Political Economy*, 21 (3), 598-642.

Schneider, F., and Enste, D. H. (2013). *The shadow economy: An international survey*. Cambridge University Press.

Schuetze, H. J. (2002). Profiles of tax non-compliance among the self-employed in Canada: 1969 to 1992. *Canadian Public Policy / Analyse de Politiques*, 28(2), 219–238.

Slemrod, J., and Weber, C. (2012). Evidence of the invisible: toward a credibility revolution in the empirical analysis of tax evasion and the informal economy. *International Tax and Public Finance*, 19, 25-53.

Trivellato, U. (2007). Protezione dei dati personali e ricerca scientifica. *Statistica*, *63*(4), 627-648.

Figure 1 - Distribution of individual adjustment factors

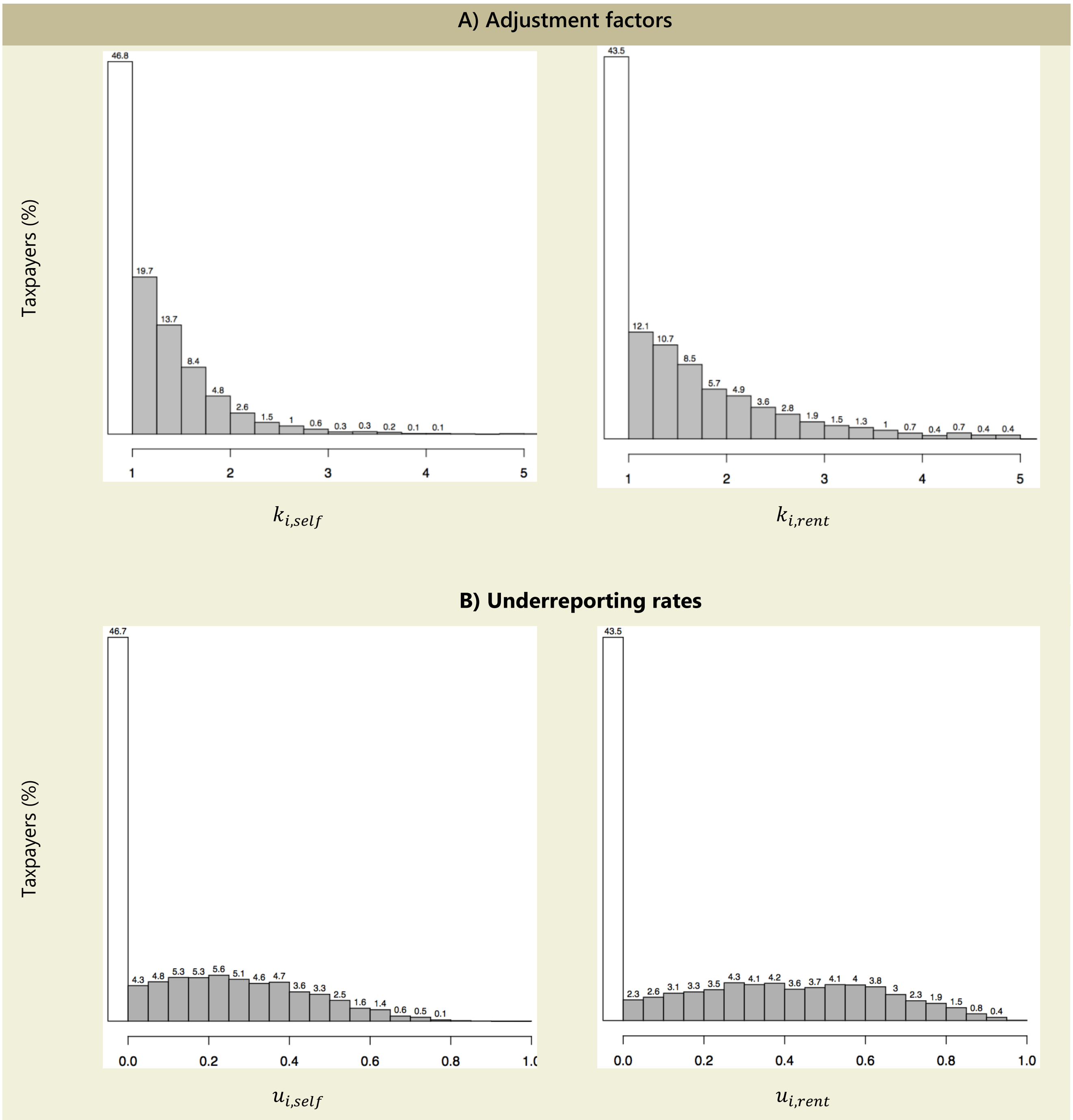
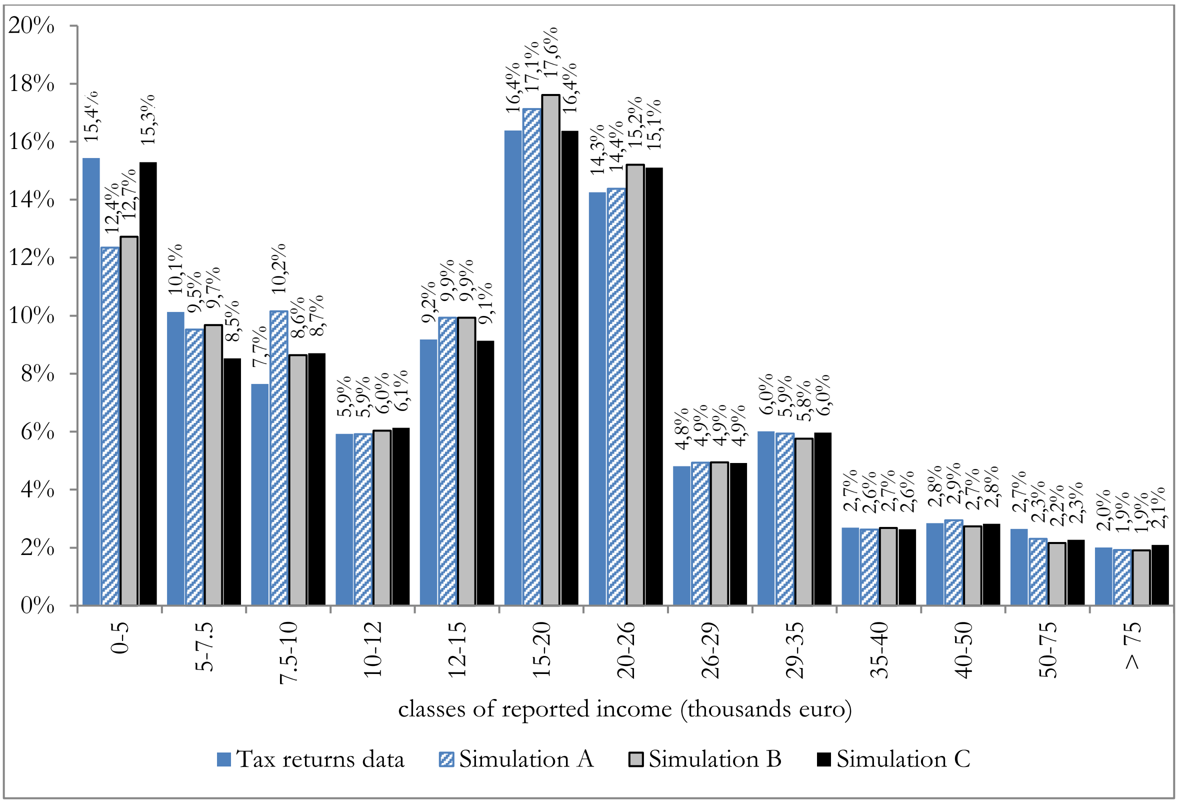


Figure 2 - Distributions of taxpayers and reported income by classes of reported income

a) Taxpayers (%)



b)Reported incomes (%)

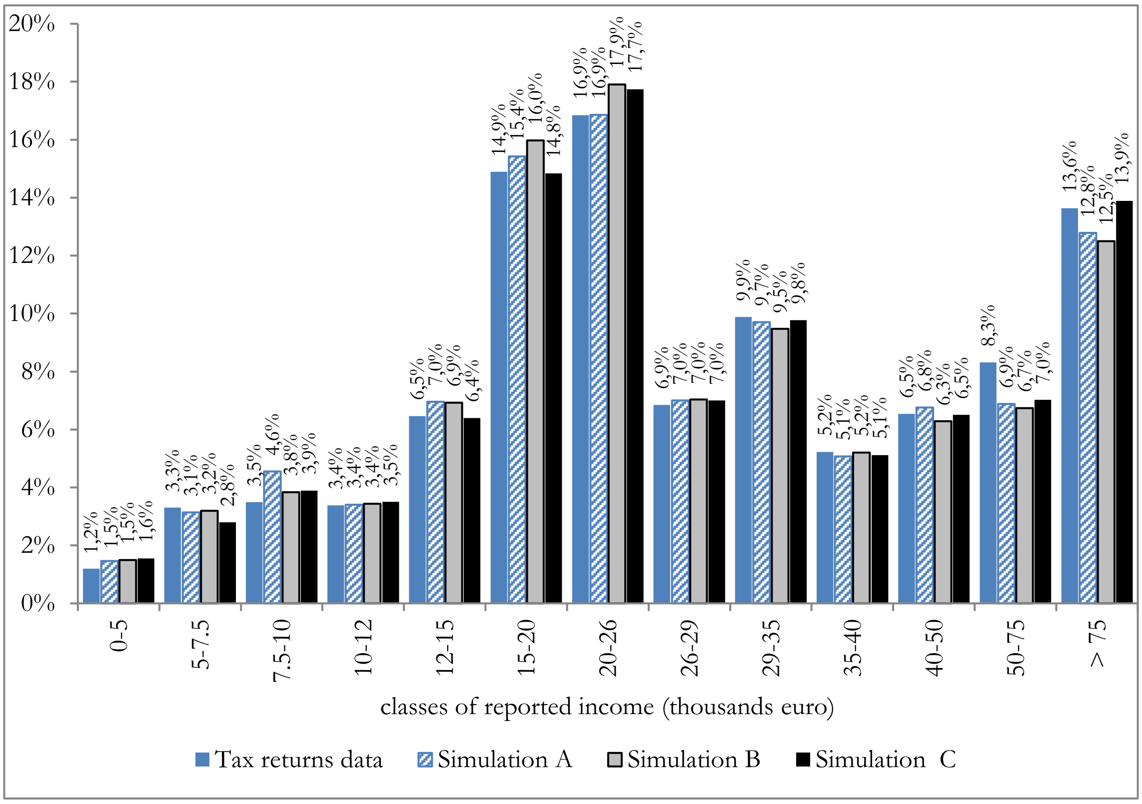
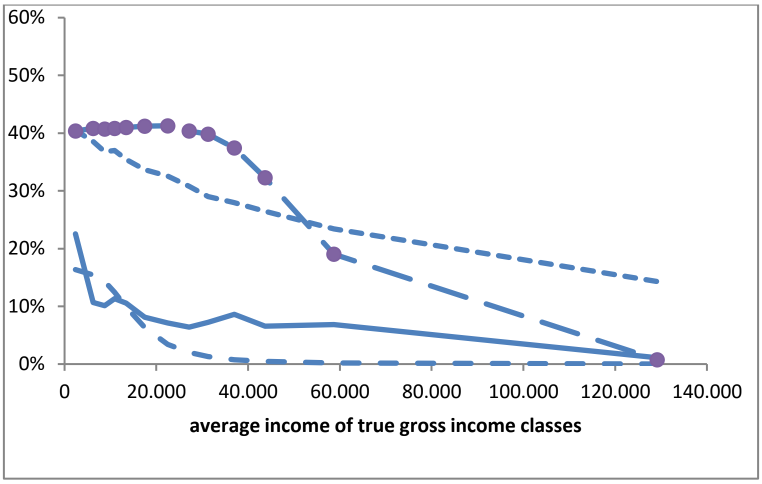
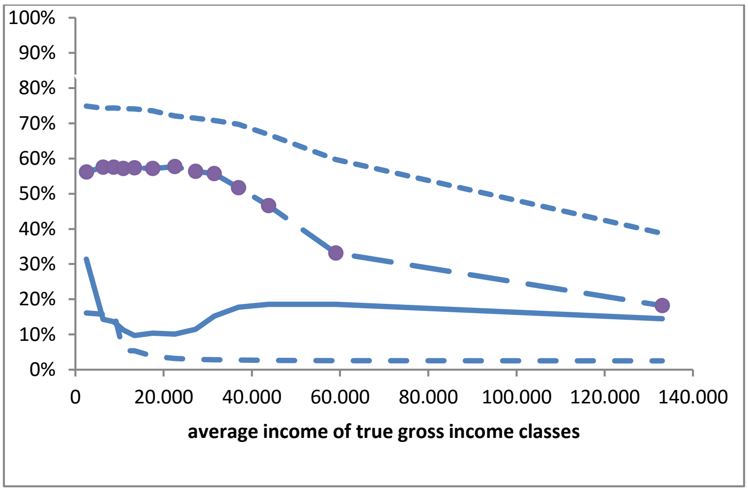


Figure 3 - Evasion rates by gross income classes

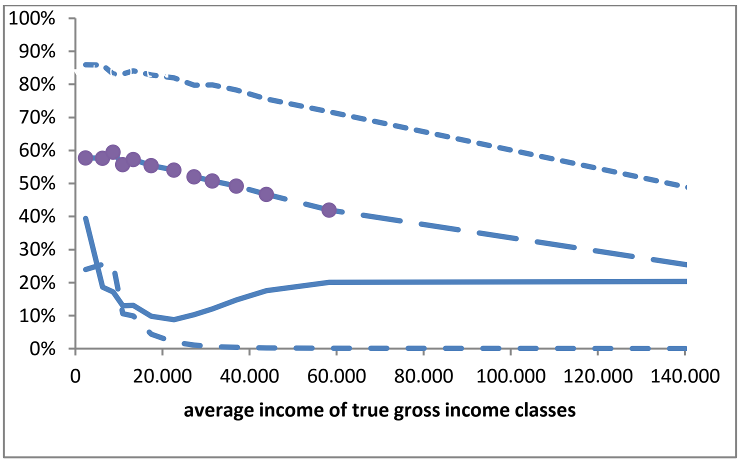
Simulation A



Simulation B



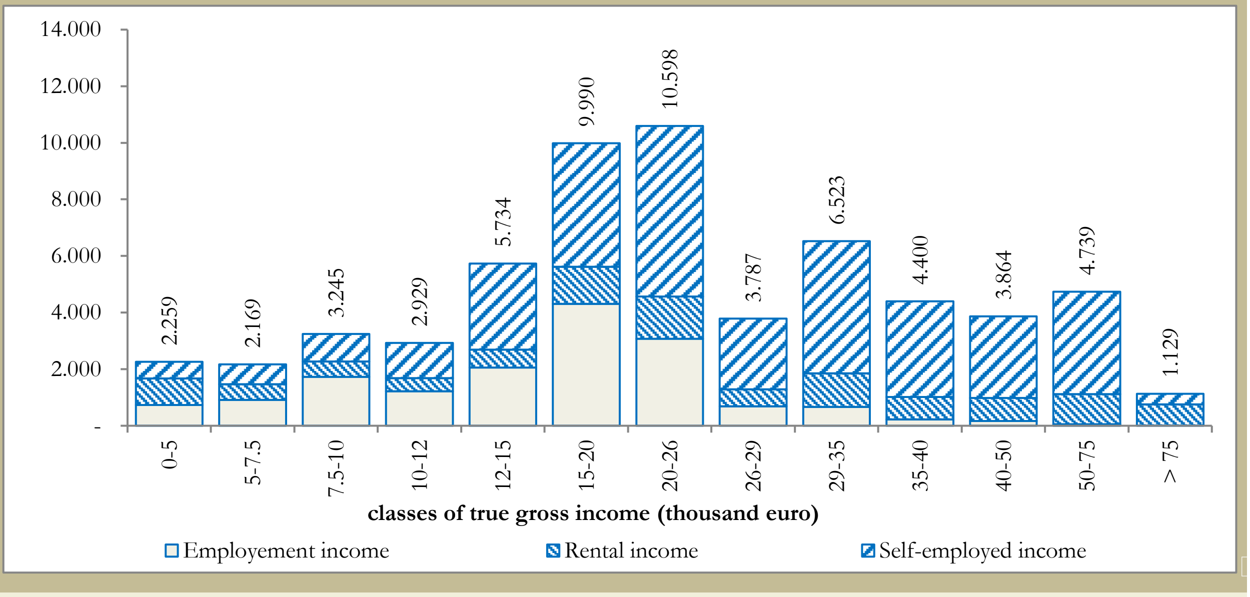
Simulation C



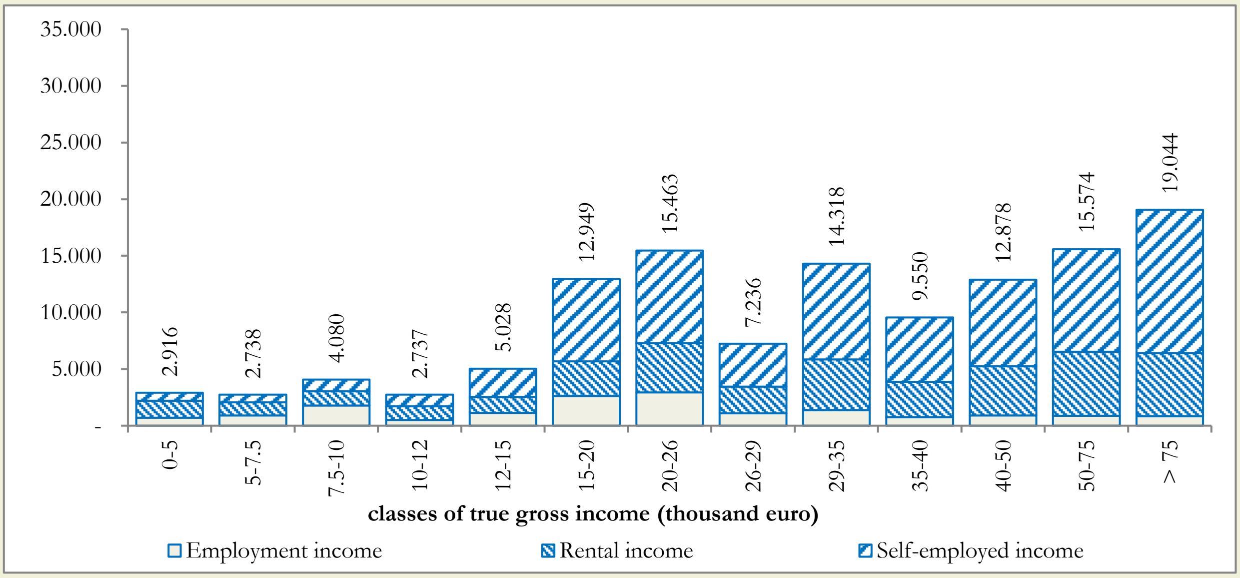
Total income   Employment income   Rental income   Self-employed income

Figure 4 - Evaded income (in millions of euro) by classes of gross incomes (thousands of euro)

Simulation A



Simulation B



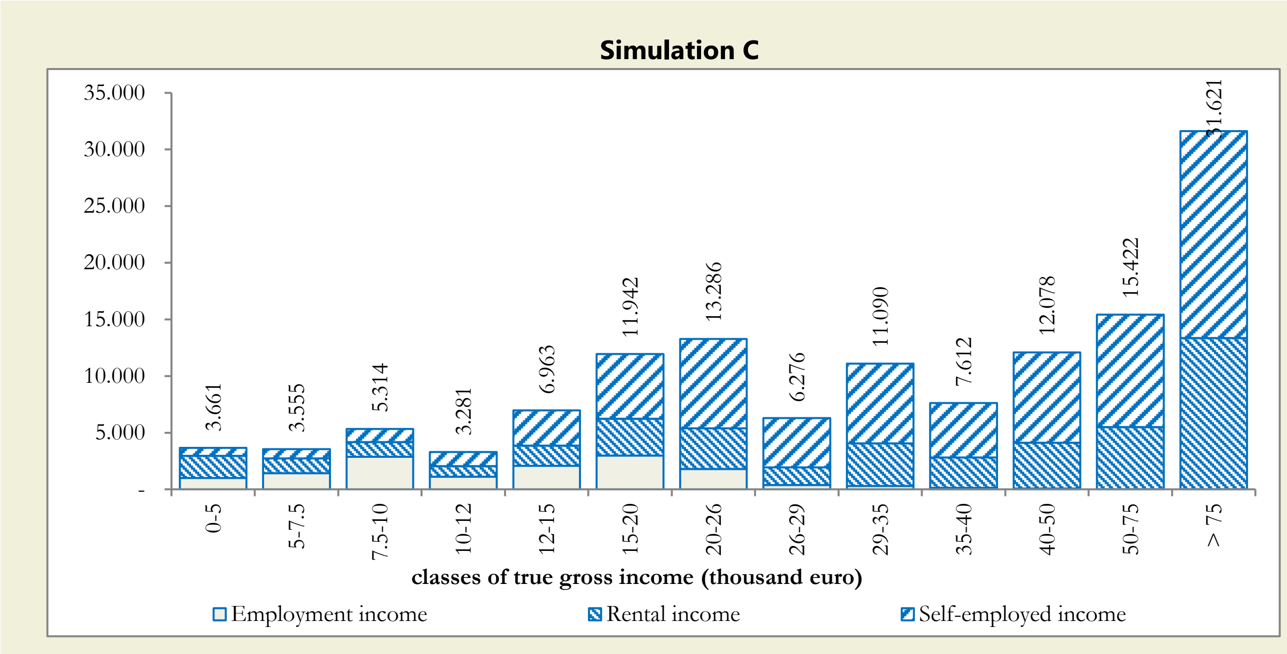
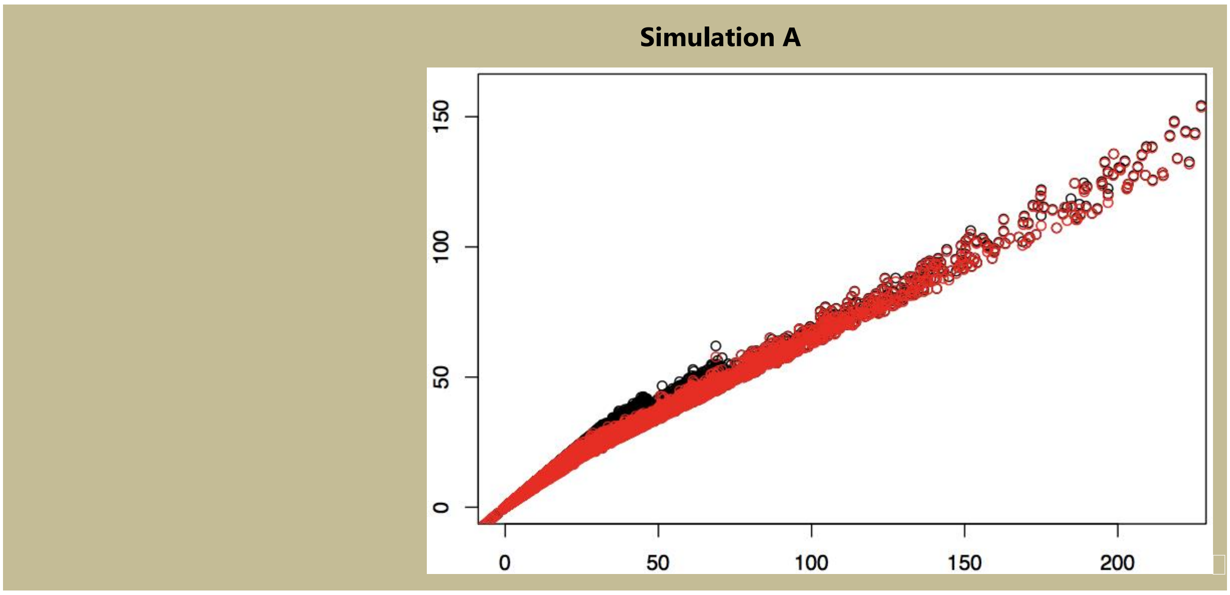


Figure 5 - Gross to net incomes tax schedules, with and without evasion



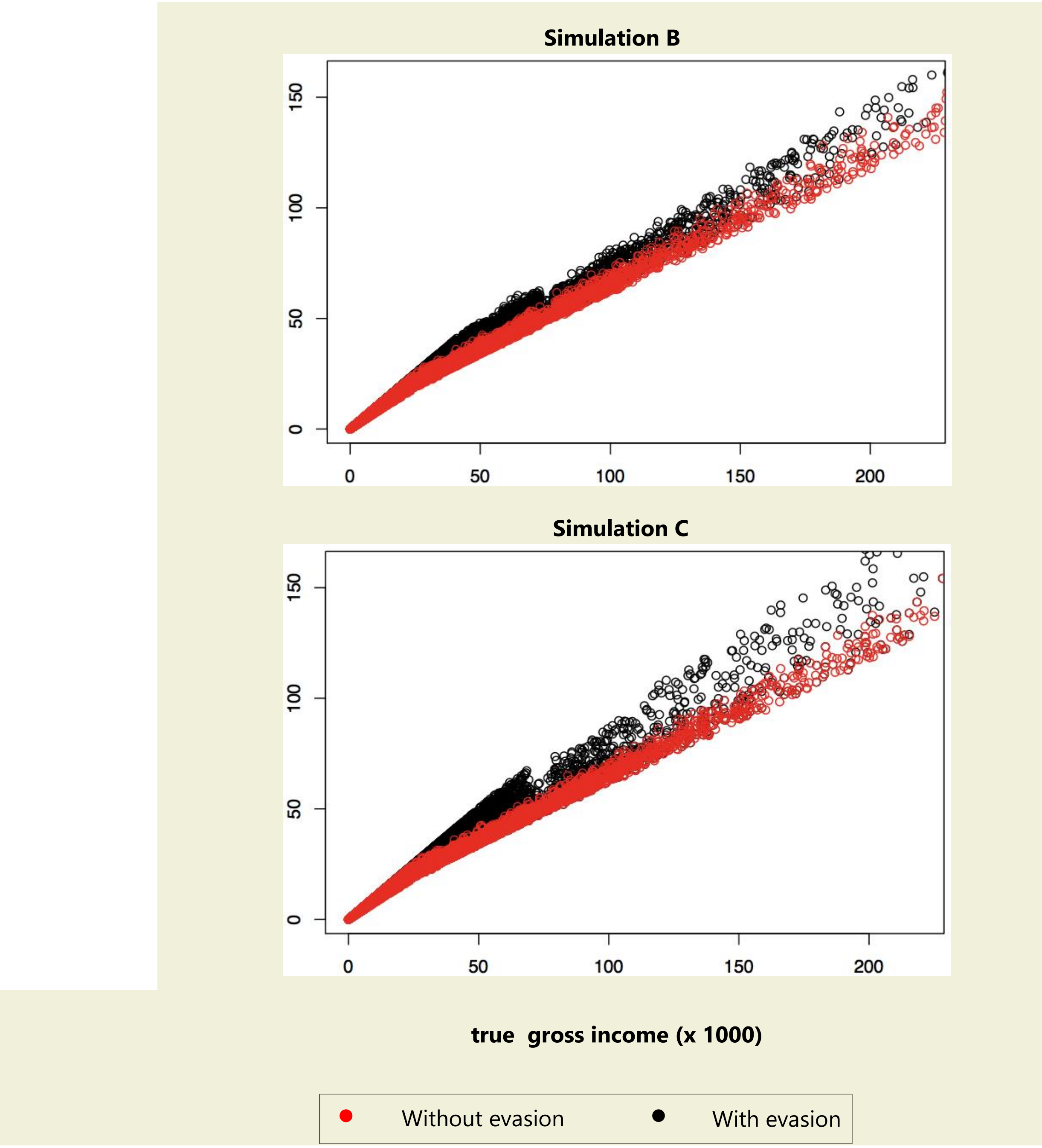

Simulation B
0
50
100
150
0
50
100
150
200
Simulation C
0
50
100
150
0
50
100
150
200
true gross income (x 1000)
Without evasion
With evasion

Figure 6 - Distribution of evasion rates by type of income (taxpayers percentages)

Table 1 – The estimated housing costs equation

Log of housing costs	**(1)** **IV**	**(2)** **Current incomes**	**(3)** **IV** $\bar{k}_{emp}$ (unre-stricted)	**(4)** **IV** (pension not only income)
Constant	4.945***	4.901***	4.946***	4.948***
	[0.066]	[0.064]	[0.066]	[0.068]
Income elasticity β_1	0.086***	0.093***	0.086***	0.083***
	[0.005]	[0.005]	[0.006]	[0.005]
Underreporting k				
Employment income			1.031***	
			[0.155]	
Self-employment income	1.298***	1.211***	1.338***	1.300***
	[0.117]	[0.134]	[0.218]	[0.116]
Rents and incomes from capital	1.773***	1.600***	1.807***	1.641***
	[0.448]	[0.395]	[0.481]	[0.426]
Household and home characteristics				
Log of number of household members	0.205***	0.202 ***	0.205***	0.195***
	[0.010]	[0.009]	[0.010]	[0.010]
Age of head of household	.006***	0.006***	.006***	0.006***
	[0.001]	[0.001]	[0.001]	[0.001]
Sq. age of household head	-.0001***	-0.0001***	-.0001***	-0.0001***
	[0.00002]	[0.00002]	[0.00002]	[0.00002]
Sex of head of household	-0.003	-0.006	-0.004	-0.012
	[0.010]	[0.010]	[0.010]	[0.010]
Proportion of males	-0.057***	-0.053 ***	-0.006***	-0.043***
	[0.015]	[0.015]	[0.015]	[0.016]
Child 0_4 (% of hh members)	-0.063	-0.067	-0.062	-0.040
	[0.039]	[0.039]	[0.040]	[0.038]
Child_5_16 (% of hh members)	-0.030	0.025	-0.030	-0.008
	[0.026]	[0.026]	[0.026]	[0.026]
Log of house size (square meters)	0.365***	0.356***	0.365***	0.374***
	[0.013]	[0.013]	[0.013]	[0.013]
Log of room number	0.017	0.016	0.015	0.008
	[0.013]	[0.013]	[0.013]	[0.014]
Housing ownership=owned	0.052***	0.052***	0.052***	0.054***
	[0.008]	[0.008]	[0.009]	[0.009]
North-east regions	-0.114***	-0.114***	-0.115***	-0.114***
	[0.010]	[0.010]	[0.010]	[0.010]
Centre Italy	-.217***	-.217***	-.217***	-0.216***
	[0.010]	[0.010]	[0.010]	[0.010]
South and isles	-0.354***	-0.353***	-0.354***	-0.337***
	[0.010]	[0.010]	[0.010]	[0.010]
Number of obs	19,004	19,043	19,004	16,692
R-squared	0.2834	0.2872	0.2834	0.2768
Adj R-squared	0.2828	0.2866	0.2828	0.2761
Root MSE	.4541634	.4533678	.4541747	.4478595
Res. dev.	23914.36	23896.69	23914.31	20536.28

Table 2 - Main aggregates of personal income tax (IRPEF) and local taxes

	Number of taxpayers [1]				Value [2]			
	Official tax returns	BETAMOD Simulation A	BETAMOD Simulation B	BETAMOD Simulation C	Official tax returns	BETAMOD Simulation A	BETAMOD Simulation B	Simulation C
Gross income	-	41,168	41,168	41,168		846,249	917,540	925,651
Evaded income	-	31,521	31,594	23,280		61,365	124,511	132,101
Reported income	41,168	41,168	41,168	41,168	792,520	793,167	793,029	793,550
		(0,0)	(0.0)	(0.0)		(0.1)	(0.1)	(0.1)
Deductions	13,374	13,799	13,726	13,664	21,746	21,728	21,751	21,726
		(3,2)	(2.6)	(2.2)		(-0.1)	(0.0)	(-0.1)
Taxable income	39,894	41,112	41,155	41,140	762,185	763,442	762,998	763,540
		(3,0)	(3.1)	(3.1)		(0.2)	(0.1)	(0.2)
Gross tax liability	39,078	41,112	41,155	41,140	205,613	204,689	204,134	206,085
		(5,2)	(5.3)	(5.2)		(-0.5)	(-0.7)	(0.2)
Tax credits	39,088	40,058	39,986	39,883	62,482	64,509	64,379	64,386
		(2,4)	(2.3)	(2.0)		(3.3)	(3.0)	(3.0)
Net tax liability	30,897	31,511	31,638	31,023	149,443	146,443	146,056	149,143
		(2,0)	(2.4)	(0.4)		(-2.0)	(-2.2)	(-0.2)
Regional income tax	30,653	31,354	31,512	30,785	8,633	8,616	8,627	8,664
		(2,3)	(-0.1)	(0.4)		(-0.2)	(0.0)	(0.3)
Municipal income tax	25,265	25,256	25,261	25,257	3,021	3,017	3,023	3,022
		(0.1)	(0.0)	(0.0)		(-0.1)	(0.0)	(0.0)

Notes: [1] thousands of persons - [2] millions of euro (in bracket % diff. from official data)

Table 3 - Evasion rates by income source and geographical area (%)

	NW	NE	C	S	ITALY
SIMULATION A					
Employment income	3.29	3.83	3.46	4.26	3.69
Pensions	0	0	0	0	0
Self-employment income	21.26	24.73	22.34	27.47	23.64
Rental income	25.89	30.19	27.25	33.54	28.96
Total income	**6.66**	**7.52**	**6.85**	**7.82**	**7.25**
SIMULATION B					
Employment income	3.74	3.86	3.79	3.97	3.84
Pensions	0	0	0	0	0
Self-employment income	36.13	37.88	36.80	39.60	37.44
Rental income	60.19	62.79	61.18	65.35	62.24
Total income	**13.64**	**13.25**	**13.60**	**13.75**	**13.69**
SIMULATION C					
Employment income	3.17	3.38	3.07	3.51	3.28
Pensions	0	0	0	0	0
Self-employment income	38.17	40.70	36.93	42.21	39.37
Rental income	63.92	68.22	61.88	70.62	65.93
Total income	**14.79**	**14.34**	**13.34**	**14.34**	**14.27**

Table 4 - Estimates of tax gaps (millions of euro)

	Simulation A		Simulation B		Simulation C	
	Net tax liability	Tax gap	Net tax liability	Tax gap	Net tax liability	Tax gap
Without evasion (theoretical)	146,443	-	146,056	-	149,143	-
With evasion on:						
- all types of income	162,927	16,484	183,517	37,460	187,715	38,571
- only employment income	150,395	3,952	150,634	4,578	151,821	2,678
- only self-employment income	155,506	9,063	166,081	20,025	170,058	20,914
- only rental income	149,803	3,360	158,669	12,612	163,889	14,746

Table 5 - Inequality and redistributive indices

Simulation A

	Without evasion (theoretical)		With evasion (actual)	
	Gini	Concentration	Gini	Concentration
True gross income	0.4142106	0.4142106	-	-
Reported income	0.4142106	0.4142106	0.4324309	0.4262497
Taxable income	0.4162223	0.4149118	0.4394466	0.4274874
Gross tax liability	0.4814227	0.4801473	0.502714	0.4914241
Net tax liability	0.6426771	0.6323645	0.6767609	0.6567132
Net income	0.3635686	0.3628381	0.3652036	0.3640206
Reynolds-Smolensky index	0.0513726		0.0501450	
Kakwani index	0.2181538		0.2425476	
Average tax rate	0.1906031		0.1713232	
Reranking effect	0.0007305		0.0011829	

Simulation C

	Without evasion (theoretical)		With evasion (actual)	
	Gini	Concentration	Gini	Concentration
True gross income	0.4350295	0.4350295	-	-
Reported income	0.4350295	0.4350295	0.4525105	0.4257737
Taxable income	0.4351068	0.4338407	0.4580904	0.4238304
Gross tax liability	0.5089212	0.5076954	0.5212436	0.4891962
Net tax liability	0.6600685	0.6508322	0.6862088	0.6466387
Net income	0.3808482	0.3801334	0.3960499	0.3942947
Reynolds-Smolensky index	0.0548961		0.0406682	
Kakwani index	0.2158027		0.2116758	
Average tax rate	0.2027941		0.1611619	
Reranking effect	0.0007149		0.0017552	

SENATO DELLA REPUBBLICA

UFFICIO VALUTAZIONE DI IMPATTO
IMPACT ASSESSMENT OFFICE

www.senato.it/ufficiovalutazioneimpatto