the review of income and wealth

Review of Income and Wealth
Series 66, Number 4, December 2020
DOI: 10.1111/roiw.12444

INCOME UNDERREPORTING AND TAX EVASION IN ITALY: ESTIMATES AND DISTRIBUTIONAL EFFECTS*

BY ANDREA ALBAREA, MICHELE BERNASCONI, ANNA MARENZI* and DINO RIZZI

Department of Economics, Ca' Foscari University of Venice

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 misreporting 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 misreporting at different rates for different income sources. The misreporting estimates are then used in the discrepancy method to correct the incomes compared with administrative registered 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 distributional impact of tax evasion.

JEL Codes: C63, D31, H26

Keywords: household surveys, income misreporting, personal income tax, tax-benefit microsimulation, tax evasion

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 on which effects of tax evasion one wants to measure.

Here 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 distributional effects of this type of tax evasion.

Pissarides and Weber (1989) developed the first method, known as the consumption-based approach. It uses micro-economic observations from consumption-expenditure surveys to estimate the consumption function for certain classes

Note: We would like to thank Roberto Casarin, Carlo Fiorio, Raffaello Seri, Francesca Zantomio and, especially, an anonymous referee and the editor for several helpful suggestions.

*Correspondence to: Anna Marenzi, Department of Economics, Ca' Foscari University of Venice, Cannaregio, 873, Venezia 30121, Italy (anna.marenzi@unive.it).

[1]Slemrod and Weber (2012) also provided a methodological review of various approaches, distinguishing between micro methods—or, as they are sometimes called, bottom-up or direct methods—and approaches based on macro-economic aggregates, which are referred to as top-down or indirect methods (see also Schneider, 2005; Giovannini, 2011; Alm and Embaye, 2013; Schneider and Enste, 2013 for a review of micro and macro studies in Italy).

of goods. After controlling for several household characteristics, the method uses differences in consumption propensities estimated for various categories of income earners to measure their tendency to misreport their incomes. 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. For example, Pissarides and Weber (1989) used employees as the reference category, while the self-employed were estimated as substantially underreporting. The method has since been applied to estimate misreporting rates in various countries and for other income categories (studies include Besim and Jenkins, 2005; Feldman and Slemrod, 2007; 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.

The consumption-based method can be used to estimate tax evasion further assuming that people behave in the survey as they do in filing their official tax-returns. This assumption can be criticized. 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 data of official tax-return registers. Typically, these comparisons show that the distributions obtained from the surveys have higher incomes than the distributions obtained from the registers, with the differences interpreted as measures of evasion. For this reason, the procedure is also called the discrepancy approach. Indeed, according to Feige (1990, p. 995), "the discrepancy approach is feasible whenever independent means exist to estimate the same conceptual entity. If one procedure for measuring a particular form of underground activity is believed to be relatively free of biases induced by the activity, while another is known to be affected by the activity, the discrepancy between the two can be used to measure the net effect of the underground activity".[2] Clearly, the method here assumes that people report their income truthfully in surveys, just as they do with 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 method is often combined with microsimulation analyses and the evasion rates estimated by the discrepancy method are then employed to compare the income distribution to counterfactual distributions simulated assuming full tax compliance in order to determine the distributional impact of tax evasion. Analyses carried out with this approach have been conducted to investigate tax evasion in Italy (e.g. Marenzi, 1996; Cannari *et al.*, 1997; Fiorio and D'Amuri, 2006; Baldini *et al.*, 2009) 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, which in addition to the tendency to underreport income may include other forms of measurement errors, e.g. due to inaccuracies, sampling errors, misclassifications (Atkinson and Brandolini, 2001).

[2]The discrepancy method can be based on various measures, including in macro studies conducted to obtain aggregate estimates of evasion via some *gap* that can be estimated (Alm, 2012). One of the most famous application is for example in the so called currency method, based on the gap between incomes and expenditures (Caridi and Passerini, 2001; Ahumada *et al.*, 2007; Ardizzi *et al.*, 2014).

The present paper integrates econometric estimates of survey misreporting in tax-benefit microsimulation analyses to estimate tax evasion in Italy. While studies employing the consumption-based method typically interpret their estimates in the context of tax evasion assuming that survey reporting behavior corresponds to the reporting for tax purposes, we depart from this by taking an agnostic view on why survey incomes may deviate from their true values - it could be tax evasion, survey measurement error or both.

Our integrated approach is illustrated in Figure 1 and discussed in details in the paper. We start by considering the possibility of income misreporting in the Italian Survey of Income and Living Conditions (IT-SILC), which is the Italian part of the

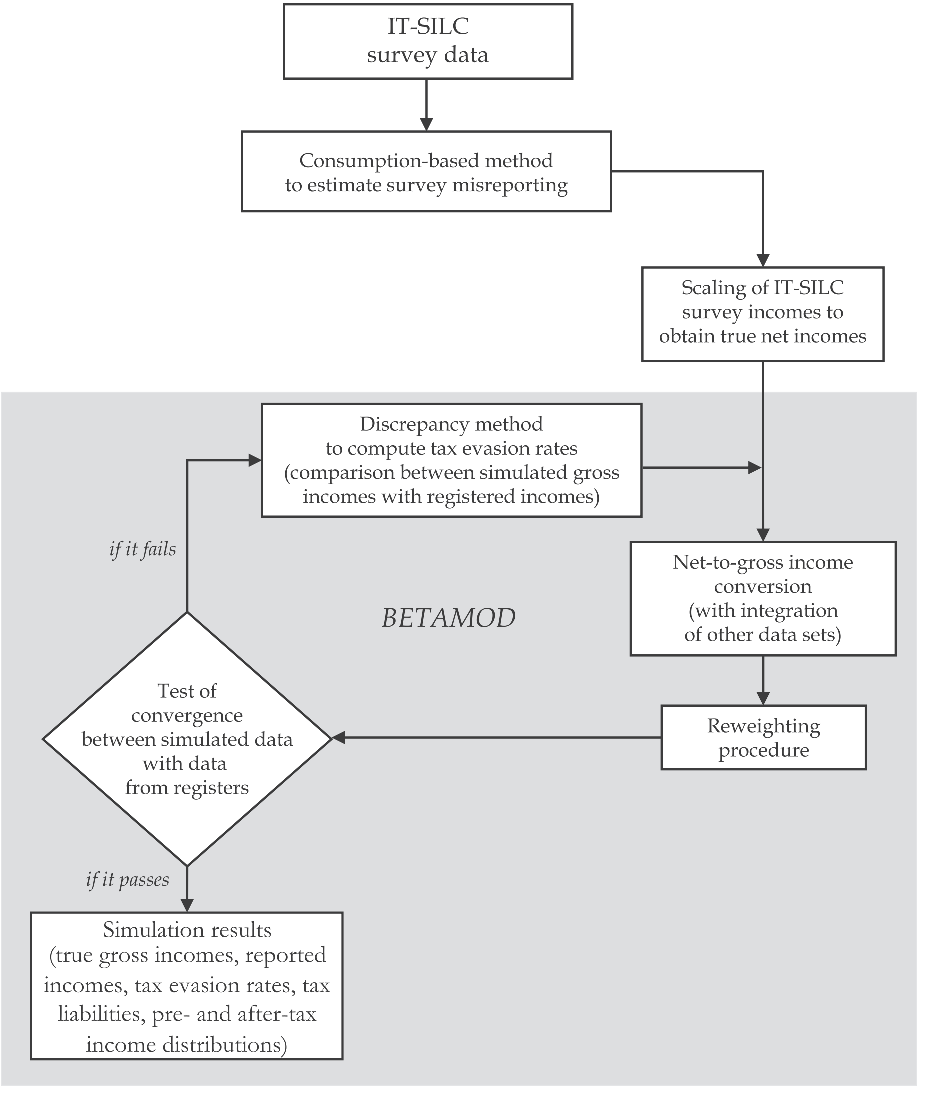

Figure 1. Integrating Survey Misreporting in Tax-Benefit Microsimulation Model BETAMOD

European Statistics on Income and Living Conditions (ISTAT, 2011). We take an econometric specification that combines Pissarides and Weber's (1989) model with Feldman and Slemrod's (2007) approach in order to estimate survey misreporting at different rates on different income sources. We use the misreporting estimates to scale the survey net incomes, which enter microsimulation analyses. These are conducted with a tax-benefit microsimulation called BETAMOD (Albarea *et al.*, 2015). BETAMOD works iteratively through various steps. In one of step it uses the discrepancy method to compare the results of the simulations with registered tax-return data to provide new estimates of PIT evasion in Italy. Thus, the consumption-based method is used here to infer only true net incomes, not tax evasion rates (i.e. income concealed for tax purposes), which are determined subsequently by comparing simulated gross incomes with official registered incomes.

We obtain several results. Previous studies that have applied the discrepancy method 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 survey misreporting in the discrepancy method improves the estimates' alignment with macro studies. With our approach we estimate an overall tax evasion rate for PIT close to 13.5 percent; without taking into account survey misreporting the estimate is about 7.2 percent, implying that the evasion captured by survey misreporting alone is about 6.3 percent. The econometric analysis confirms that self-employment incomes and rental incomes are substantially misreported in the IT-SILC survey. We also test for, but do not find, misreporting of employment incomes in the survey. The microsimulation analysis improves our ability to study the distributional profile of tax evasion. The evasion rates on specific income sources are generally decreasing in the various income components. 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 organized into two main parts with several subsections. We start from the consumption-based approach to estimate income misreporting in the survey. Then we consider the microsimulation-discrepancy approach and integrate the two methods. A concluding section discusses our approach further by considering the problem of the availability of data to conduct studies that seek to analyze tax evasion.

2. The Consumption-based Approach to Estimating Income-Misreporting in Surveys

2.1. *Methodology*

The consumption-based approach to estimating income misreporting in surveys 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 misreport income.

Pissarides and Weber (1989; PW hereafter) and Feldman and Slemrod (2007; FS hereafter) provide two models of income misreporting. PW developed a model on two types of households, employees and the self-employed. They focus on misreporting by the self-employed and assume that the degree of misreporting refers to all of their income. FS took a model of multiple income sources in which a household can misreport incomes in different amounts if they come from different sources. They also assume that a given income source is misreported in the same proportion by all households.

We consider a model similar to FS. They, however, 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. The log-linear Engel curve for a consumption of good C (food in PW) is:

$$\ln C_i = \beta_0 + \beta_1 \ln y_i^T + X_i \beta_2 + u_i, \tag{1}$$

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.

PW assumed a model with two types of households, employees and the self-employed, distinguished on the basis of their main source of income. They assume that only self-employed misreport.[3] Here we maintain more generality and, following FS, use a model in which the total household income is the sum of several components:

$$y_i^T = \sum_j y_{ij}^T, \tag{2}$$

where y_{ij}^T is the component from source j for household i.

Income misreporting occurs when the income reported in a survey by household i from source j, namely y_{ij}^R, is different than the true income. Following FS, we take that a given income source is misreported in the same proportion by all taxpayers. In particular, let $\bar{k}_j \neq 1$ denote an adjustment factor that measures the extent of misreporting by any household i on income source j. Thus, we assume the following relationships between the true income and the reported one:

$$y_{ij}^T = \bar{k}_j y_{ij}^R, \tag{3}$$

with $\bar{k}_j$ greater than 1 indicating underreporting.[4]

[4]Alternatively, in possibly more realistic specifications, one could model household propensities to misreport as $k_{ij} = \bar{k}_j e^{v_{ij}}$, where v_{ij} is a term for household i idiosyncratic propensity to misreport income source j. One would however then need to assume specific restrictions to estimate the model. For example, PW estimate a model starting from a similar specification for k_{ij}, but restricting it to apply only to self-employed households. In a set-up of multiple income sources the restrictions to estimate the model are in general more severe.

[3]In PW the definition of the main source of income for a household is based on an exogenously assumed threshold. They define self-employed households as households with reported income from self-employment of at least 25 percent of total reported income. The threshold is then checked by controlling that the estimates do not change significantly moving the threshold in a comparable range.

By substituting equations (2) and (3) in (1), we obtain:

$$\ln C_i = \beta_0 + \beta_1 \ln \left[\sum_j \bar{k}_j y_{ij}^R \right] + X_i \beta_2 + \mathrm{u}_i \tag{4}$$

Equation (4) is indeed very similar to the model studied by FS.[5] The identification strategy is based on the assumption that one income source, the reference income source (say $\bar{k}_1 = 1$), is correctly reported (more on this point below). The model can be estimated by nonlinear least squares. Its estimation provides directly the adjustment factors $\bar{k}_j$ from which one obtains the misreporting rates for source j, given by:

$$\bar{u}_j = 1 - \frac{1}{\bar{k}_j} \tag{5}$$

2.2. *Survey Data or Official Tax Data?*

The original study by PW was conducted using survey data (mainly, the 1982 wave of the Family Expenditure Survey). They found that the level of income misreporting by self-employed households in the UK averaged 33 percent (37 percent among blue-collar households and 29 percent 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 misreporting can be directly interpreted as non-compliance. However, charitable contributions reported for tax deductions or other data from tax registers could themselves be altered or reported untruthfully. For example, some taxpayers may exaggerate charitable contributions to benefit from tax deductions. Others may simply forget to report part 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 Canada (Schuetze, 2002), Sweden (Engström and Hagen, 2017), the US (Hurst *et al.*, 2014), the UK (Cabral *et al.*, 2014), North Cyprus (Ekici and Besim, 2016), and several

[5]There are nevertheless differences between equation [4] and the model estimated by FS. The main ones are due to the fact that FS use official tax income data as a dependent variable and take charitable contributions reported for tax deductions as dependent variables. Below we will discuss in more details advantages and disadvantages of using survey versus official data.

others quoted in, e.g. Paulus (2015a, 2015b). The studies have provided consistent evidence of underreporting of self-employed households compared to employed households (in a range between 15 percent and 40 percent), 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 linking data about consumption and other information from surveys with data from official tax-return incomes. Linking survey data with official tax-return data is appealing because it offers the possibility of interpreting income misreporting 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 analyses for Finland (Johansson, 2005), Sweden (Engström and Hagen, 2017), Estonia (Paulus, 2015a, 2015b). Paulus (2015b) compares the misreporting estimated from survey-reported income data with official tax-return data. The analysis confirms that the extent of misreporting is higher for official tax-return incomes than it is for survey-reported incomes. 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 misreporting 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. 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, household savings, debt, means-tested benefits, 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.[6]

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-misreporting, but it includes a rich battery of expenditures for running a household that have been tested as a way to estimate misreporting (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.

[6]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. This type of control will improve the quality of data for research purposes (more on this point in the conclusion).

We take as our main reference income category the aggregate of pensions and unemployment benefits and other state benefits that are subject to withholding. We estimate potential misreporting 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. Most previous analyses based on survey data have taken employee households as the reference income households. A problem with this choice, however, is that employees may themselves misreport income. We seek to verify this possibility in the data set.

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 misreporting 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 our investigation is based on a full data set using the 2011 cross-sectional wave of IT-SILC, which includes 19,043 households with positive home-utility expenditures.

2.4. *Estimation and Results*

We estimate several specifications of the equation (4) taking expenditures for home utilities as the dependent variable and using the nonlinear least-square estimation. As previously indicated, our identification strategy is based on the assumption that pensions are correctly reported so that we fix $\bar{k}_{pension} = \bar{k}_1 = 1$. We notice that this allows to identify the parameters $\bar{k}_j$ for all other incomes $j \neq 1$ in equation (4). Nevertheless, precision of the estimates increases with the level of heterogeneity in the composition of households' incomes including for what concern the presence of households with only one source of income. In Table A.1 of the online Appendix A, we report the composition of households' incomes in our sample, which confirms that the composition of households' incomes is quite heterogeneous.

The literature has considered various income measures as independent variables in estimating misreporting. FS estimated misreporting using current income sources, while PW estimated misreporting using instrumental variables to reduce the effect of measurement errors and transitory components of income.

In the specification (1) of Table 1 we estimated the model using instrumental variables. Suitable instruments respecting identifying restrictions have been used for employment incomes and self-employment incomes. 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. Diagnostics checks for the first stage based on R^2 and F-statistics

TABLE 1
THE ESTIMATED HOUSING COSTS EQUATION

	(1)	(2)	(3)	(4)
			IV	IV
Log of Housing Costs	IV	Current Incomes	$\bar{k}_{emp}$ (Unrestricted)	(Pension not Only Income)
Constant	2.475***	1.950***	2.506***	2.900***
	[0.423]	[0.477]	[0.436]	[0.435]
Income elasticity β_1	0.110***	0.123***	0.109***	0.106***
	[0.005]	[0.005]	[0.005]	[0.005]
Underreporting k				
Employment income	-	-	1.040	-
			[0.142]	
Self-employment income	1.282++	1.207+	1.332	1.281++
	[0.105]	[0.081]	[0.200]	[0.108]
Rents and incomes from capital	1.709+	1.315	1.748+	1.560
	[0.347]	[0.259]	[0.375]	[0.340]
Household characteristics				
Log number of household members if average household	0.150***	0.148***	0.150***	0.152***
age ≤ 30	[0.009]	[0.009]	[0.009]	[0.009]
Log number of household members if average household	0.188***	0.186***	0.188***	0.186***
30 < age ≤ 60	[0.009]	[0.009]	[0.009]	[0.009]
Log number of household members if average household	0.228***	0.226***	0.229***	0.209***
age ≥ 60	[0.015]	[0.015]	[0.015]	[0.016]
Age of head of household	0.925***	1.113***	0.909***	0.695**
	[0.224]	[0.253]	[0.231]	[0.231]
Sq. age of household head	−0.113***	−0.139***	−0.110***	−0.080**
	[0.029]	[0.033]	[0.031]	[0.031]
Sex of household head	−0.005	−0.011	−0.006	−0.010
	[0.009]	[0.009]	[0.009]	[0.009]
Economic sector of household head: Primary	0.029	0.036	0.028	0.025
	[0.025]	[0.025]	[0.025]	[0.025]
Economic sector of household head: Tertiary	−0.013	−0.002	−0.014	−0.011
	[0.009]	[0.008]	[0.009]	[0.009]

TABLE 1 (CONTINUED)

	(1)	(2)	(3)	(4)
			IV	IV
Log of Housing Costs	IV	Current Incomes	$\bar{k}_{emp}$ (Unrestricted)	(Pension not Only Income)
Proportion of males	−0.047***	−0.041***	−0.047***	−0.036**
	[0.014]	[0.014]	[0.014]	[0.015]
Home characteristics				
Log of house size (square meters)	0.387***	0.378***	0.387***	0.392***
	[0.013]	[0.013]	[0.013]	[0.014]
Log of room number	0.033*	0.031*	0.033*	0.027
	[0.013]	[0.013]	[0.013]	[0.014]
Housing ownership = owned	0.061***	0.061***	0.061***	0.061***
	[0.009]	[0.008]	[0.008]	[0.009]
Semi-detached house	0.021*	0.018*	0.021*	0.015
	[0.009]	[0.009]	[0.009]	[0.010]
Apartment in a building with less than 10 dwelling	0.042***	0.039***	0.042***	0.033**
	[0.010]	[0.010]	[0.010]	[0.011]
Apartment in a building with more than 10 dwelling	0.168***	0.163***	0.168***	0.154***
	[0.010]	[0.010]	[0.010]	[0.011]
Centralized heating	0.276***	0.272***	0.276***	0.275***
	[0.012]	[0.012]	[0.012]	[0.013]
Year of construction of the main residence after 1995	0.013	0.008	0.013	0.015
	[0.011]	[0.011]	[0.011]	[0.011]
Year of construction of the main residence before 1960	−0.016*	−0.016*	−0.016*	−0.006
	[0.008]	[0.008]	[0.008]	[0.008]
Number of obs	18,949	18,994	18,949	16,642
R-squared	0.2783	0.2823	0.2783	0.2729
Adj R-squared	0.2775	0.2815	0.2774	0.2720
Root MSE	0.4544	0.4536	0.4544	0.4477
Res. dev.	23,859.43	23,849.77	23,859.33	20,460.75

Notes: Stars (*) denote significantly different from 0 for all coefficients (***p < 0.001, **p < 0.01, *p < 0.05); plus (+) denote significantly different from 1 for the coefficients of underreporting k (++p < 0.01, +p < 0.05). Standard errors are in square brackets.

tell us that the models fit the data relatively well.[7] Various other covariates, including additional households characteristics (like household head's gender, age, age squared, economic sector of occupation, number of household members in various age groups) and homes characteristics (like number of rooms, size in m^2, type of ownership, type of house, year of construction) are in the estimation of equation (4). The covariates for the characteristics are generally statistically significant. We also remark that in the regression we haven't included geographical variables already used as instruments.

The IV estimates of model (1) are consistent with the hypothesis of misreporting. The adjustment factors on self-employment income and rent income are, respectively, $\hat{\bar{k}}_{self} = 1.282$ and $\hat{\bar{k}}_{rent} = 1.709$. Both coefficients are significantly greater than 1 (the former at 1% statistical level and the second at the 5% level). The corresponding misreporting rates are $\hat{\bar{u}}_{self} = 22$ percent for self-employment and $\hat{\bar{u}}_{rent} = 42$ percent for rents.[8]

In model (2) of Table 1 the same specification is estimated using current incomes. The estimates of the misreporting coefficients are a bit lower than those with IV, especially for the adjustment factor on rent income, which becomes not significantly different from 1. This 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 (0.110 and 0. 123, 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.

Model (3) in Table 1 adds the possibility of misreporting on employment income. The point estimate on employment income is $\hat{\bar{k}}_{empl} = 1.040$, which is not significantly different from 1 at statistical level. Thus, the model rejects the hypothesis that employment income is misreported.

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 generally robust to the exclusion.

Overall, we therefore interpret the results speaking in favor of specification (1) as our best fitting model. This is because model (3) does not reject $\hat{\bar{k}}_{empl} = 1$ for employment income.

[7]The method of instrumental variables is standard in models linear in variables and parameters. Diagnostics tests are studied in special nonlinear models and not available for the general nonlinear case (see e.g. Stock *et al.* 2002). F-test and R^2 are given here as standard checks for the overall significance of the first stage regression. The statistics are: $R^2 = 0.435$ and F = 155.98 ($p \simeq 0.0$) in the regression for employment incomes, and $R^2 = 0.278$ and F = 33.10 ($p \simeq 0.0$) in the regression for self-employment incomes.

[8]The coefficients $\hat{\bar{k}}_{self}$ and $\hat{\bar{k}}_{rent}$ measure the extent of misreporting at the intensive margin. It could also be interesting to test for possible effects at the extensive margin. For example, FS use dummies to separate taxpayers filling the different income tax schedules (even if they reported 0 income in the schedules) from the taxpayers not filling the schedules and found that merely filling the income schedules is associated in official tax return data with a higher income. The same procedure is not possible with survey data.

3. Evasion Rates and Microsimulation Analysis with Correction for Survey Misreporting

Income misreporting in a survey may include only a part of the non-compliance committed by taxpayers. We integrate income misreporting 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 income data from official tax registers. However, surveys and registered data do not contain the same information, so the method requires adjustments in order to ensure that the income variables and the population underlying the income distributions from the surveys and from the registers are consistently defined and comparable. In the surveys the respondents are typically asked for their disposable income, whereas data from tax registers usually comes in the form of tables of reported gross income and taxes.[9] Moreover, registered 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 (discussion and references in D'Amuri and Fiorio, 2006; Marino and Zizza, 2008).

The literature that has applied the discrepancy method has followed various procedures to address the above issues. Our application of the discrepancy approach is based on microsimulation analysis[10] and is part of a microsimulation model called BETAMOD.

3.1. *The Microsimulation Model BETAMOD*

As anticipated in the Introduction, BETAMOD is a model for the Italian PIT (IRPEF, with regional and municipal surtaxes). It works through various steps and modules in an iterated process (see Figure 1). A full description of the model is in Albarea *et al.* (2015).

Simulations begin with the conversion of the individual net incomes from IT-SILC, here appropriately rescaled to account for survey misreporting (see below for details), into gross incomes.[11] Since IRPEF is reported on an individual basis, BETAMOD uses the household information from IT-SILC 2011 to conduct simulations at the individual level. It is also worthwhile to remark that since the IT-SILC survey does not cover all individual and household information relevant to perform a precise net-to-gross income conversion, BETAMOD is enriched with additional information from two other population surveys. Mainly, the 2010 Survey on Households Income and Wealth (SHIW) released by the Bank of Italy (2012), is used to compute cadastral values and tax relief for imputation of payments of insurance premiums and other home-refurbishments expenditures, and the 2013

[9]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 net-to-gross conversion of the incomes (Betti *et al.*, 2011).

[10]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 Sutherland, 1991; Bourguignon and Spadaro, 2006; Immervoll *et al.*, 2007; Figari *et al.*, 2015).

[11]In our empirical analysis "true gross income" refers to the tax base of PIT and does not include the social insurance contributions.

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.

A standard reweighting procedure is also performed in BETAMOD to adjust the IT-SILC probability-sampling weights so that simulation results are consistent with the tax-registered data. Mainly, while IT-SILC weights are designed to produce population totals from the national statistics, we adjust them to achieve consistency with tax register data as well, so that the model estimates are reconciled with both the entire populations and the taxpayers counts from the register.

The net-to-gross income conversion is based on an algorithm similar to those used in the literature (e.g. in Immervoll and O'Donoghue, 2001). It estimates the gross incomes of individual taxpayers by applying to the net incomes of the survey the tax rules determined on the basis of individual and households characteristics, and taking as given the individual tax evasion rates (zero in the first-round simulation). In essence, the net-to-gross income conversion is obtained from the following transformation:

$$Y_i = \frac{y_i}{1 - \tau_{D_i}\left(1 - e_i\right)} \tag{6}$$

where Y_i is the PIT taxable gross income of individual i to be obtained, y_i is the true total disposable income of unit i based on IT-SILC, τ_{D_i} is the households' average tax rate simulated by BETAMOD, and e_i is the household imputed tax evasion rate. As ordinarily defined, the latter is given by $e_i = \left(Y_i - D_i\right) / Y_i$, where D_i is the reported PIT income.[12]

Round-specific convergence measures are then assessed in term of consistency with registered data and the model is iterated until convergence is achieved; that is, the iterations stop when the distributions of reported levels of income simulated by the BETAMOD do not differ significantly from the distributions of official reported incomes D_i, at both the aggregate level and the subgroup level, with the latter defined by main source of income and geographic area. When it stops, the model generates a battery of individual level variables, which include true gross incomes, tax evasion rates, reported incomes, tax relevant expenditures, calibration weights, the pre- and post-tax income distributions.

When convergence fails, the model is iterated. The iterations start by producing new estimates of the tax evasion rates using the discrepancy method. Evasion rates are computed as the percentage differences between the simulated true gross

[12]We emphasize that the above procedure of estimating the PIT evasion rate is valid even considering possible evasion of social contributions. To see it, we clarify that in Italy social contributions are paid before personal income taxes and are deductible from the PIT tax basis. This means that with $\tilde{Y}_i$ denoting the gross household's income *before* both social contributions *and* PIT, the taxpayer's disposable income is $y_i = \tilde{Y}_i\left[1 - t_{cs}\left(1 - \tilde{e}_i\right)\right]\left[1 - \tau_{D_i}\left(1 - e_i\right)\right]$, where t_{cs} is the social contribution rate and $\tilde{e}_i$ the evasion rate on social contributions (which may be equal or different from the PIT evasion rate e_i). Then, given the deductibility of social contributions, the PIT tax basis is $Y_i = \tilde{Y}_i\left[1 - t_{cs}\left(1 - \tilde{e}_i\right)\right]$, which substituted in y_i above gives equation [6] for the relationships between Y_i, e_i and y_i.

incomes and the incomes reported in the registers.[13] As register incomes are provided at semi-aggregate level by four main income sources (employment income, pensions, self-employment, rental from immovable property) and four geographical areas (northwest, northeast, central, south and isles), the model firstly produces a 4x4 matrix of average tax evasion rates by income type and geographical area.

BETAMOD then further estimates, for each area-by-income-type stratum, a distribution of tax evasion rates by thirteen gross income classes. In particular, since for all the taxpayers, registered data are also available for classes of total annual reported income, BETAMOD uses a procedure that expands each of the 16 cells of the 4x4 matrix of the average tax evasion rates by income type and geographical area, into 16 profiles of thirteen tax evasion rates by gross income classes. It does so by applying a numerical algorithm that, given the individual gross incomes of the round-specific simulation and the 4x4 average tax evasion rates by income type and geographical area, minimizes the distance between the distributions of total reported income from the simulation and the register. The final result is thus a (4x4)×13-dimension matrix of evasion rates by main income source, geographical area, and class of true gross income level.[14]

A further important point is that 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 distributional impact of tax evasion, and ignoring them may lead to a substantial underestimation of the regressive impact of tax evasion.

Two final remarks are in order. Firstly, we note that the discrepancy approach based on the comparison with registered data allows us to detect only the compliance behaviors of the taxpayers in the register, and not of nonfilers. These may also represent a part of tax evasion, which however is not easy to evaluate in the microsimulation. Secondly and partly related, the approach is used here to estimate the evasion of the personal income tax, but it could also be applied to estimate the evasion of social security contributions. In order to do that, however, one would need access to registered data on payroll taxes, which aren't unfortunately available in a comparable form.

3.2. *Integrating the Consumption-based Approach into BETAMOD*

The evidence obtained in Section 2.4 from the consumption-based method rejects the assumption that income is truthfully reported in the IT-SILC survey data. Accordingly, net incomes of the micro-units from IT-SILC that represent the input of BETAMOD have been corrected using the misreporting rates estimated by the consumption-based approach. In fact, we conducted two simulations.

[13]In other words, in the first round simulation ($s = 1$), it is assumed $e_i = 0$ so that $Y_i = D_i$. From round-simulation $s = 2$ onwards the round-imputed evasion rates are given by $e_i^s = \frac{Y_i^{s-1} - D_i}{Y_i^{s-1}}$, with iterations stopping when the distributions of simulated reported incomes, at both aggregate and subgroup levels, are equal to the distributions of official reported incomes D_i.

[14]The classes of annual gross income (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, 50-75, and >75.

The first simulation (simulation A) is run as the benchmark and is conducted with the original net incomes of the micro-units from IT-SILC. The second simulation (simulation B) corrects the input data to take account for survey misreporting. In particular, simulation B is conducted multiplying the various income components j of individual i by the mean adjustment factor $\hat{\bar{k}}_j$, so the income of micro-unit i used as input by BETAMOD is:

$$y_i^B = \sum_j \hat{\bar{k}}_j y_{ij}, \tag{7}$$

where, as we recall, $\hat{\bar{k}}_{self} = 1.282$ and $\hat{\bar{k}}_{rent} = 1.709$.

Therefore, with the comparison between simulations A and B we can take account of how the estimates of the tax evasion rates obtained by the discrepancy-method are affected by the individuals' propensities to misreport in the survey and also check how the scale of misreporting in the surveys may differ from the scale of cheating to the tax authorities.

3.3. *Results of the Simulations and Measures of Tax Evasion*

Table 2 provides aggregate quantifications for the main components of the Irpef and local taxes and compares them with official data. All simulations provide evidence of the consistency of the simulations with the official tax-return data:[15] 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 misreporting, estimates that on the aggregate slightly more than €61.3 billion in gross income escapes the tax authorities, corresponding to an evasion rate of 7.2 percent. Correcting for misreporting in simulations B raises the estimates to €121.2 billion in total gross income evaded, respectively, corresponding to evasion rates of 13.3 percent.

These numbers attest to the relevance of the problem of misreporting in the survey to the ability to quantify the dimension of evasion. In particular, an estimated evasion rate of around 7 percent is low compared to that obtained by other studies and methodologies (e.g. Marino and Zizza, 2008). The rates estimated by simulation 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. In this respect we also notice that the difference between simulation B and simulation A on the other side can be used to estimate

[15]More discussion and evidence on internal and external validity of the microsimulation procedure and of the reweighting are provided in the online Appendix B.

TABLE 2
MAIN AGGREGATES OF PERSONAL INCOME TAX (IRPEF) AND LOCAL TAXES

	Number of Taxpayers[a]			Value[b]		
		BETAMOD			BETAMOD	
	Official Tax Returns	Simulation A	Simulation B	Official Tax Returns	Simulation A	Simulation B
Gross income	-	41,168	41,168		854,532	913,433
Mean					20,757	22,188
Standard error					154.58	173.27
Evaded income	-	31,521	31,590		61,365	121,150
Mean					1,490	2,943
Standard error					18.97	47.78
Reported income	41,168	41,168	41,168	792,520	793,167	792,283
		(0.0)	(0.0)		(0.1)	(−0.0)
Mean					19,267	19,245
Standard error					152.09	149.46
Deductions	13,374	13,799	13,736	21,746	21,728	21,775
		(3.2)	(2.7)		(−0.1)	(0.1)
Taxable income	39,894	41,112	41,156	762,185	763,442	762,230
		(3.0)	(3.2)		(0.2)	(0.0)
Gross tax liability	39,078	41,112	41,156	205,613	204,690	203,750
		(5.2)	(5.3)		(−0.4)	(−0.9)
Tax credits	39,088	40,058	39,984	62,482	64,510	64,387
		(2.4)	(2.3)		(3.2)	(3.0)
Net tax liability	30,897	31,511	31,666	149,443	146,443	145,637
		(2.0)	(2.5)		(−2.0)	(−2.5)
Regional income tax	30,653	31,354	31,536	8,633	8,617	8,614
		(2.3)	(2.9)		(−0.2)	(−0.2)
Municipal income tax	25,265	25,256	25,262	3,021	3,017	3,019
		(0.1)	(−0.0)		(−0.1)	(−0.1)

Notes: [a]Thousands of persons.
[b]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	6.67	7.52	6.85	7.82	7.18
	(0.177)	(0.196)	(0.181)	(0.212)	(0.097)
Simulation B					
Employment income	3.78	3.90	3,83	4.02	3.88
Pensions	0	0	0	0	0
Self-employment income	35.70	37.47	36.47	39.21	37.04
Rental income	59.18	61.77	60.32	64.34	61.26
Total income	13.30	12.96	13.34	13.41	13.26
	(0.325)	(0.301)	(0.346)	(0.398)	(0.174)

Note: Standard errors in brackets.

the evasion that is captured in the survey misreporting alone. This indicates that survey misreporting can capture about half of the whole evasion (6.3%).

For true gross incomes, evaded incomes, and reported incomes, Table 2 also reports means and standard errors to reflect sample variability (Goedemé *et al.*, 2013). Differences between the means estimated across simulations are statically significant for gross incomes and evaded incomes (p-values approximately 0), whereas consistently with the design of BETAMOD they are not statistically significant for reported incomes.

Table 3 reports the average evasion rates by income source and geographic area and documents the impact of correcting for misreporting in the estimation of evasion rates. The estimates from all simulations confirm that tax evasion on employment income (between 3.7-3.9 percent in the two 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 percent on self-employment and 29.0 percent on rental income), increase in simulation B, that corrects for misreporting, to 37.0 percent for self-employment and to 61.3 percent rental income. The percentages in Table 3 also reveal some differences among geographic areas: in particular, all simulations identify the south of Italy as having systematically higher evasion rates on the individual income components, followed by the northeast. Differences across regions are nevertheless not very large. This may be explained recalling that the estimations are here based on comparisons with registered taxpayers, whose compliance behavior may vary relatively little between regions. Larger regional differences are more typically related with activities escaping any form of taxation, including those due to very small businesses of the informal sector, non-filers and illegal activities (e.g. D'Attoma, 2017, and references therein).

The overall tax evasion rates depend on the income composition. The income shares are shown in Table 4. The figures indicate that employment income accounts for the largest income share, corresponding in all simulations to around half of total income.

TABLE 4
Share of Income by Income Source and Geographical Area (%)

	NW	NE	C	S	Italy
Simulation A					
Employment income	49.6	52.4	50.1	51.1	50.7
Pensions	27.5	25.7	27.7	29.2	27.6
Self-employment income	18.0	18.4	17.4	14.8	17.2
Rental income	4.9	3.4	4.8	4.9	4.5
Total income	100.0	100.0	100.0	100.0	100.0
Simulation B					
Employment income	46.1	49.1	46.6	47.7	47.3
Pensions	25.9	24.3	26.1	27.4	25.9
Self-employment income	20.8	21.5	20.2	17.5	20.0
Rental income	7.2	5.0	7.2	7.4	6.8
Total income	100.0	100.0	100.0	100.0	100.0

TABLE 5
Estimates of Tax Gaps (Millions of Euro)

	Simulation A		Simulation B	
	Net Tax Liability	Tax Gap	Net Tax Liability	Tax Gap
With evasion	146,443	-	145,636	-
Without evasion on:				
All types of income	162,927	16,484	182,016	36,380
Only employment income	150,395	3,952	150,256	4,619
Only self-employment income	155,506	9,063	165,243	19,606
Only rental income	149,803	3,360	157,526	11,889

Table 5 shows the losses in tax revenues (tax gap) 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 and €36.4 billion in simulation B. In both simulations the greater part of the tax gap is caused by evasion on self-employment income, which in the simulations B amounts to €19.6. The tax gap on rental income is also much higher in simulation B (€11.9) than in simulation A, which does not correct for misreporting (about €3.3 billion).

Finally, the tax gaps estimated on self-employment in simulations B is consistent with the tax gap that a recent official report (MEF, 2016) obtained for the same fiscal. In particular, the official tax gap on self-employment incomes estimated by the report using the discrepancy method at macro level (based on the comparison between income reported on tax returns and income measured in the national income accounts) is €20.1 billion.[16] The similarity between the estimate

[16]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. The estimate in this case was obtained by imputing taxes due, but not paid taxes, on the irregular jobs estimated directly by the Italian Central Institute of Statistics. We also notice that, unfortunately, in no case the official document reported the estimates of the total incomes associated with the various tax gaps.

here and those from MEF is another indication of the importance of correcting for income misreporting in the discrepancy approach at the micro level.

3.4. *Distributional Effects of Tax Evasion*

An advantage of microsimulation analysis is that it permits one to study the distributive effects of evasion. Figure 2 reports the average tax evasion rates by income source and gross income class obtained with the simulations. A feature common to both simulations A and B is that the evasion rates computed for employment incomes, self-employment incomes and rental incomes all have a negative gradient. This is consistent with studies that have shown evasion rates generally

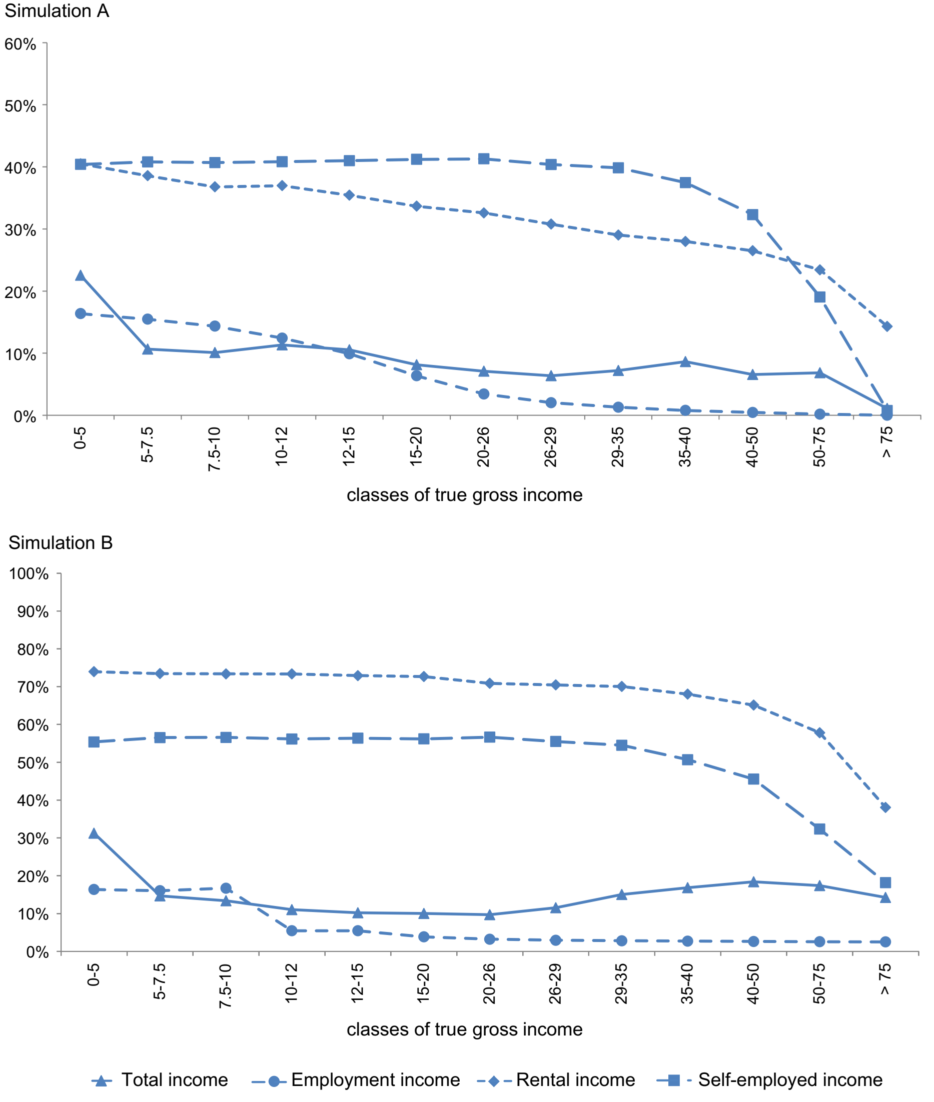

Figure 2. Average Tax Evasion Rates by Income Classes [Colour figure can be viewed at wileyonlinelibrary.com]

decreasing in the three components of income (Bernasconi and Marenzi, 1999; Fiorio and D'Amuri, 2006). However, unlike previous studies, the present analysis finds in both 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 in simulation A (similar to previous works), while in simulation B the evasion rate for total income is firstly decreasing and then slightly.

These impacts on total income are due to the composition effects illustrated in Figure 3. The figure shows the total amount of unreported income-by-income

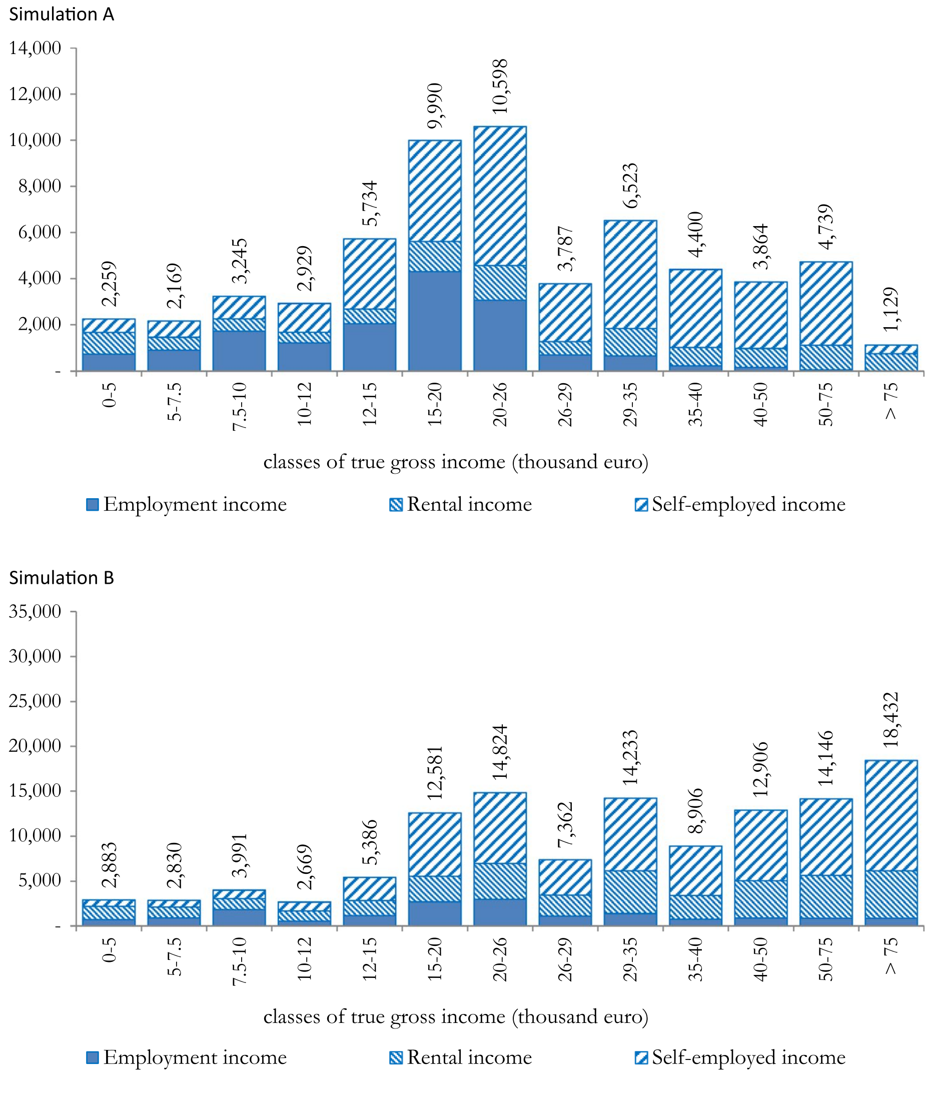

Figure 3. Evaded Income (in Millions of Euro) by Classes of Gross Incomes (Thousands of Euro) [Colour figure can be viewed at wileyonlinelibrary.com]

class and income source in the two 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 simulation B 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, modifying the redistributive effect of the tax schedule, which can change its progressivity impact and have horizontal inequity effects and reranking effects. Table 6 reports a set of standard inequality indices to evaluate the redistributive impact of tax evasion in simulation B (similar computations for Simulation A are reported in the online Appendix D). 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 (0.430 versus 0.419). However, evasion increases inequality in the distribution of after-tax incomes. The concentration index for the distribution of net incomes with evasion (0.378) is higher than the index for the theoretical distribution without evasion (0.368); 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.

Figure 4 presents the distributions of the evasion rate across income sources for the two simulations. The comparison shows the effect of taking account of misreporting in simulation B that increases with respect to simulation A the estimates of the tax evasion rates on self-employment income and rents and, consequently, those for total incomes.

Finally, Table 7 shows some evasion indices by type of income based on the evasion profiles of Figure 5 (Rizzi, 2017). Concentration indices of evasion rates (C_e) are higher for employment income (25.05 percent), whereas rents show the lowest concentration index (11.69 percent). The composite index of evasion $E=\bar{e}\left[1+C_e\right]$

TABLE 6
Inequality and Redistributive Indices

	Simulation B			
	Without Evasion (Theoretical)		With Evasion (Actual)	
	Gini	Concentration	Gini	Concentration
True gross income	0.4192	0.4192	-	-
Reported income	0.4192	0.4192	0.4300	0.4108
Taxable income	0.4187	0.4176	0.4367	0.4092
Gross tax liability	0.4888	0.4876	0.4987	0.4727
Net tax liability	0.6398	0.6303	0.6693	0.6338
Net income	0.3674	0.3667	0.3800	0.3784
Reynolds-Smolensky index	0.0525		0.0407	
Kakwani index	0.2111		0.2147	
Average tax rate	0.1993		0.1594	
Reranking effect	0.0007		0.0016	

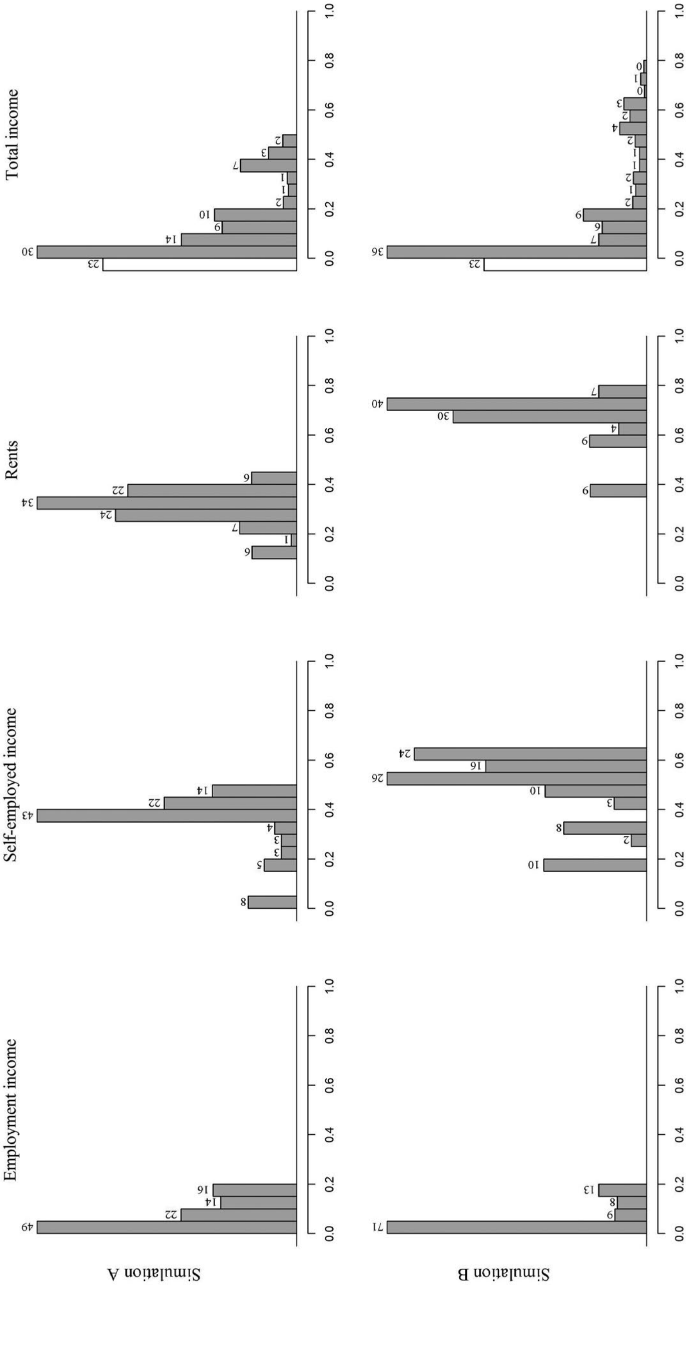

Figure 4. Distribution of Tax Evasion Rates by Type of Income

TABLE 7
Indices of Tax Evasion by Type of Income

Indices		Employed Incomes (%)	Self-Employed Incomes (%)	Rents (%)	All Incomes (%)[a]
Intensity of undeclared incomes (average evasion rate)	$\bar{e}$	3.88	37.04	61.26	13.26
Concentration index of evasion rates	C_e	25.05	25.21	11.69	64.55
Composite index of evasion	$E=\bar{e}\,[1+C_e]$	4.85	46.38	68.42	21.82

Note: [a]Including pensions.

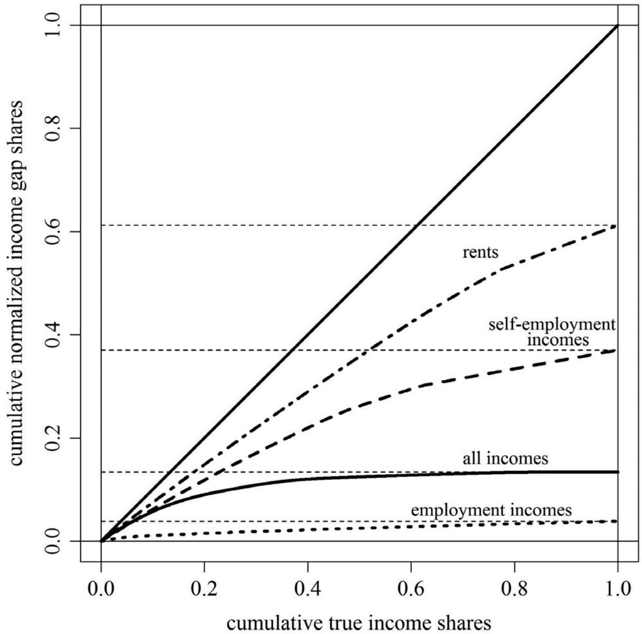

Figure 5. Evasion Profiles by Type of Income

is equal to 22.02 percent for all incomes (including pensions); employed incomes show the lowest values (4.85 percent) and rents confirm the highest value of evasion with 68.42 percent. The profiles in Figure 5 confirm visually the ranking, the distribution of evasion rates and the relative importance of evasion by types of income.

4. Conclusion

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

We have proposed an approach that integrates the consumption-based method of estimating income misreporting in surveys with a microsimulation-based discrepancy analysis to determine evasion rates of the personal income tax in Italy. We have used the consumption-based method to estimate income misreporting of self-employment incomes and rental incomes from capital and immovable properties. We have not found signs of misreporting of employment income at a statistically significant level. Using the discrepancy method, we have found that there is a substantial amount of tax non-compliance that occurs in addition to income misreporting, and we have used micro data corrected for misreporting to estimate the distributional profile of tax evasion.

We have focused on the personal income tax in Italy, due to various reasons including availability of data. The quality of the data used in the empirical investigations on tax evasion is very important. 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 administrative data in the context of the EU-SILC (Jäntti *et al.*, 2013). There are benefits and costs for using administrative data, and the extent to which registered data are actually used in practice varies widely across countries. This also depends on the fact that few countries have yet a practice to link registered data of different administrations and surveys data.

The likely best system would use administrative data to complement data collected through surveys, rather than as a substitute. Datasets that combine survey data with administrative records could greatly benefit the study of tax evasion, particularly when incomes and information from different sources are linked at individual level. First of all, integration of different sources of data could permit to study evasion with respect to the whole system of taxation, analyze for example whether compliance behavior changes with respect to the type of taxes and/or the administration collecting the taxes, and possibly even help to detect forms of no-filing behaviors.

Econometric approaches like the consumption-based method could benefit by improved possibilities of identification, for example permitting to better identify misreporting for the purpose of tax evasion—including in analyses of income-misreporting among employees (e.g. as in Paulus, 2015a)—and separate it from other types of measurement errors.

The microsimulation discrepancy method and the integrated approach proposed here could also benefit: for deriving gross incomes and double-checking for any tax evasion remaining even after the grossing-up procedure; for computing the tax gap that is due to tax evasion with the same accurate microsimulation model used for tax and benefits; and for analyzing the distributional impact of tax evasion, along with the possibility of informing the microsimulation analysis of the people propensities to misreport survey incomes.

References

Ahumada, H., F. Alvaredo, and A. Canavese, "The Monetary Method and the Size of the Shadow Economy: A Critical Assessment," *Review of Income and Wealth*, 53, 363–71, 2007.

Albarea, A., M. Bernasconi, C. Di Novi, A. Marenzi, D. Rizzi, and F. Zantomio, "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, 2015.

Alm, J., "Measuring, Explaining, and Controlling Tax Evasion: Lessons from Theory, Experiments, and Field Studies," *International Tax and Public Finance*, 19, 54–77, 2012.

Alm, J. and A. Embaye, "Using Dynamic Panel Methods to Estimate Shadow Economies Around the World, 1984–2006," *Public Finance Review*, 41, 510–43, 2013.

Ardizzi, G., C. Petraglia, M. Piacenza, and G. Turati, "Measuring the Underground Economy with the Currency Demand Approach: A Reinterpretation of the Methodology, With an Application to Italy," *Review of Income and Wealth*, 60, 747–72, 2014.

Atkinson, A. B. and A. Brandolini, "Promise and Pitfalls in the Use of 'Secondary' Data-Sets: Income Inequality in OECD Countries as a Case Study," *Journal of Economic Literature*, 39, 771–99, 2001.

Baldini, M., P. Bosi, and M. Lalla, "Tax Evasion and Misreporting in Income Tax Returns and Household Income Surveys," *Politica Economica*, XXV, 333–48, 2009.

Bank of Italy, *SHIW, Survey on Household Income and Wealth in 2010*, Supplements to the Statistical Bulletin, 6, Rome, 2012.

Bernasconi, M. and A. Marenzi, “Gli effetti redistributivi dell’evasione fiscale in Italia,” *Ricerche quantitative per la politica economica 1997*, Convegno CIDE-SADIBA, Bank of Italy, 1999.
Besim, M. and G. P. Jenkins, “Tax Compliance: When Do Employees Behave like the Self-Employed?” *Applied Economics*, 37, 1201–8, 2005.
Betti, G., G. Donatiello, and V. Verma, “The Siena Microsimulation Model (sm2) for Net-Gross Conversion of EU-SILC Income Variables,” *International Journal of Microsimulation*, 4, 35–53, 2011.
Bourguignon, F. and A. Spadaro, “Microsimulation as a Tool for Evaluating Redistribution Policies,” *Journal of economic Inequality*, 4, 77–106, 2006.
Cabral, A. C. G., C. Kotsogiannis, and G. Myles, “Self-Employment Underreporting in Great Britain: Who and How Much?,” Tax Administration Research Center Discussion paper 010-14, 2014.
Cannari, L., V. Ceriani, and G. D’Alessio, “Il recupero degli imponibili sottratti a tassazione,” *Ricerche quantitative per la politica economica 1995*. Convegno CIDE-SADIBA, Bank of Italy, 1997.
Caridi, P. and P. Passerini, “The Underground Economy, the Demand for Currency Approach and the Analysis of Discrepancies: Some Recent European Experience,” *Review of Income and Wealth*, 47, 239–50, 2001.
D’Amuri, F. and C. Fiorio, “Grossing-up and Validation Issues in an Italian Tax-Benefit Microsimulation Model,” Econpubblica WP, No. 117, Bocconi University, Milan, 2006.
D’Attoma, J., “Divided Nation: The North-South Cleavage in Italian Tax Compliance,” *Polity*, 49, 69–99, 2017.
Ekici, T. and M. Besim, “A Measure of the Shadow Economy in a Small Economy: Evidence from Household-Level Expenditure Patterns,” *Review of Income and Wealth*, 62, 145–60, 2016.
Engström, P. and J. Hagen, “Income Underreporting Among the Self-Employed: A Permanent Income Approach,” *European Economic Review*, 92, 92–109, 2017.
Feige, E. L., “Defining and Estimating Underground and Informal Economies: The New Institutional Economics Approach,” *World Development*, 18, 989–1002, 1990.
Feldman, N. and J. Slemrod, “Estimating Tax Noncompliance with Evidence from Unaudited Tax Returns,” *The Economic Journal*, 117(518), 327–52, 2007.
Figari, F., M. Iacovou, A. J. Skew, and H. Sutherland, “Approximations to the Truth: Comparing Survey and Microsimulation Approaches to Measuring Income for Social Indicators,” *Social Indicators Research*, 105, 387–407, 2012.
Figari, F., A. Paulus, and H. Sutherland. “Microsimulation and Policy Analysis,” in A. B. Atkinson and F. Bourguignon (eds), *Handbook of Income Distribution*, Elsevier, 2B, 2141–2210, 2015.
Fiorio, C. V. and F. D’Amuri, “Tax Evasion in Italy: An Analysis Using a Tax-Benefit Microsimulation Model,” *The IUP Journal of Public Finance*, 4, 19–37, 2006.
Giovannini, E., *Economia non osservata e flussi finanziari. Rapporto finale*, Ministry of Economy and Finance, Rome, 2011.
Goedemé, T., K. Van den Bosch, L. Salanauskaite, and G. Verbist, “Testing the Statistical Significance of Microsimulation Results: A Plea,” *International Journal of Microsimulation*, 6, 50–77, 2013.
Hurst, E., G. Li, and B. Pugsley, “Are Household Surveys like Tax Forms: Evidence from Income Underreporting of the Self-Employed,” *Review of Economics & Statistics*, 96, 19–33, 2014.
Immervoll, H., and C. O’Donoghue, “Imputation of Gross Amounts from Net Incomes in Household Surveys: An Application Using EUROMOD,” EUROMOD Working Paper Series, EM1/01. University of Essex, 2001.
Immervoll, H., H. J. Kleven, C. T. Kreiner, and E. Saez, “Welfare Reform in European Countries: A Microsimulation Analysis,” *The Economic Journal*, 117, 1–44, 2007.
Jäntti, M., V. M. Törmälehto, and E. Marlier, *The Use of Registers in the Context of EU-SILC: Challenges and Opportunities*. Eurostat ed. Publications Office, Luxembourg, 2013.
ISTAT, *IT-SILC, Statistics on Income and Living Conditions*, 2011. Available at: https://www.istat.it/it/archivio/5663.
______, *Multiscopo Survey on Health Conditions and the Use of Health Services 2012/2013*, 2014. Available at: http://www.istat.it/it/archivio/7740.
Johansson, E., “An Estimate of Self-Employment Income Underreporting in Finland,” *Nordic Journal of Political Economy*, 31, 99–109, 2005.
Marenzi, A., “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, 1996.
Marino, M. R., and R. Zizza, *L'evasione dell'Irpef: una stima per tipologia di contribuente*, Banca d'Italia, Servizio Studi di Struttura economica e finanziaria, 2008.
Matsaganis, M., D. Benedek, M. Flevotomou, O. Lelkes, D. Mantovani, and S. Nienadowska. “Distributional Implications of Income Tax Evasion in Greece, Hungary and Italy,” Munich Personal RePEc Archive (MPRA) Paper No. 21465, Munich University Library, 2010.

MEF, *Relazione sull'economia non osservata e sull'evasione fiscale e contributiva*, Ministry of Economy and Finance, Rome, 2016.
Paulus, A. "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, 2015a.
______, "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, 2015b.
Pissarides, C. A., and G. Weber, "An Expenditure-Based Estimate of Britain's Black Economy," *Journal of Public Economics*, 39, 17–32, 1989.
Rizzi, D., "Tax Evasion Indices and Profiles," *Public Finance Review*, 45, 771–91, 2017.
Schneider, F., "Shadow Economies Around the World: What Do We Really Know?," *European Journal of Political Economy*, 21, 598–642, 2005.
Schneider, F. and D. H. Enste, *The Shadow Economy: An International Survey*, Cambridge University Press, Cambridge, 2013.
Schuetze, H. J., "Profiles of Tax Non-Compliance Among the Self-Employed in Canada: 1969 to 1992," *Canadian Public Policy / Analyse de Politiques*, 28, 219–38, 2002.
Slemrod, J. and C. Weber, "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, 2012.
Stock, J. H., J. H. Wright, and M. Yogo, "A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments," *Journal of Business & Economic Statistics*, 20, 518–29, 2002.
Sutherland, H., "Constructing a Tax-Benefit Model: What Advice Can One Give?," *Review of Income and Wealth*, 37, 199–219, 1991.

Supporting Information

Additional supporting information may be found in the online version of this article at the publisher's web site:

Appendix A: Composition of households' incomes.
Table A.1. Composition of households' incomes.
Appendix B: External and internal validity of simulations with BETAMOD.
Figure B.1: Taxpayers distributions and reported incomes by classes of annual reported income.
Figure B.2: Average reported and gross incomes from official tax returns and simulations
Appendix C: Effect of re-weighting.
Table C.1: Number of taxpayers and values by geographical area with reweighting and with IT-SILC weights.
Table C.2: Tax evasion rates (percentages) with EU-SILC weights.
Appendix D: Inequality and redistributive indices in simulation A.
Table D.1: Inequality and redistributive indices.