MEF Dipartimento delle Finanze

Working Paper

Size, heterogeneity and distributional effects of self-employment income tax evasion in Italy

DF WP n.8
October 2020

Martina Bazzoli, Paolo Di Caro, Francesco Figari, Carlo V. Fiorio and Marco Manzo

DF Working Papers, 2017

MEF Ministero dell'Economia e delle Finanze
Dipartimento delle Finanze
Via dei Normanni, 5
00184 Roma

Email df.wpapers@finanze.it
http://www.finanze.gov.it/

ISSN electronic edition 2610-8887

Papers can be downloaded free of charge from the MEF website:
http://www.finanze.it/opencms/it/il-dipartimento/documentazione/Collana-di-lavori-e-di-ricerca/

Size, heterogeneity and distributional effects of self-employment income tax evasion in Italy

Martina Bazzoli[*], Paolo Di Caro[†], Francesco Figari[‡], Carlo V. Fiorio[§] and Marco Manzo[¶]

October 19, 2020

Abstract

We measure tax evasion in Italy by estimating a food expenditure equation that disentangles households with prevalent income from self-employment, which is self-declared, from those with mostly third-party reported income. By using a novel dataset that links the 2013 Italian Household Budget Survey with individual tax records over a period of 7 years, we reduce measurement error by a great extent. We also depart from the usual constant share of underreporting, showing that underreporting heterogeneity among self-employed is significant, and is larger for singles and for college-educated households. We show that self-employed workers in Italy exhibit a similar attitude to tax evasion as those in other developed countries. Therefore, we point to the structure of the economy for an explanation of why aggregate tax evasion in Italy is larger than in other developed countries.

The estimated heterogeneity of underreporting behavior of households combined with the use of a tax-benefit microsimulation model have allowed us to shed light on the distributional effects of income tax evasion, showing that almost 73% of the missing revenue is attributable to taxpayers at the top of the income distribution.

[*]Irvapp-FBK

[†]Italian Ministry of Economy and Finance

[‡]University of Insubria, ISER University of Essex, CeRP and Dondena Bocconi University

[§]University of Milan, Irvapp-FBK and Dondena. Corresp. author: carlo.fiorio@unimi.it

[¶]Italian Ministry of Economy and Finance.

We wish to thank Erich Battistin, Mustafa Besim, Vieri Ceriani, Fernando Di Nicola, Christos Kotsogiannis, Jamie McGuire, Giorgio Mongelli, Maria Teresa Monteduro, Alari Paulus, Stefano Pisani, Federico Revelli and Frank Van Driessche for helpful comments. We are grateful to the participants at the HMRC International Tax Gap Conference (London), 6th Shadow Economy Conference (Trento), the XXX SIEP Conference (Padua) and seminar participants at EU-JRC (Seville), the TAIEX-SRSS tax gap workshop (Rome), the OECD tax gap CoP, the 10th Annual IRS/TPC Joint Research Conference on Tax Administration (Washington) and at the University of Milan. We also thank SOGEI s.p.a. for the support with the data. We use EUROMOD version I2.0 and we were financially supported by the European Union Easi programme (2014-2020). This paper is part of a research project developed at the Department of Finance of the Italian Ministry of Economy and Finance. The views expressed in this paper are those of the authors and do not reflect those of the institutions of affiliation.

1 Introduction

The presence of underreported income, deriving from tax evasion and noncompliance activities, challenges the sustainability of public finances, distorts the distribution of the tax burden among citizens, affects the fair and efficient functioning of the economy, and raises vertical and horizontal equity concerns in modern tax-benefit systems (Andreoni et al., 1998; Alm, 1999; Slemrod and Yitzhaki, 2002). There is increasing evidence that the share of underreported income is large all over the world. According to recent estimates for selected developed countries, the total tax gap– which is a proxy widely used for measuring tax noncompliance – is about 6% in the United Kingdom, 12% in Canada, 15% in the United States, and 22% in Italy, with a relevant share due to self-employment.[1]

Measuring personal income tax evasion is a daunting task because those who choose to cheat the government are careful to hide their fraud to avoid fines and other penalties. The gold standard for measuring tax evasion are stratified random audits, such as those conducted periodically by the United States tax authority Compliance Measurement Program/Service National Research Program (TCMP/NRP) or in few other countries (e.g. Denmark, see Kleven et al., 2011). Although these audits are ideal for defining an optimal audit strategy and they provide estimates of unreported income used in national accounts, they are costly, they require the direct involvement of the tax authority, and they rely on the ability and the selection of examiners (for an analysis of the amounts of tax evasion that goes undetected in IRS random audits, see Feinstein, 1991). Field experiments also provide a promising avenue of research to solve major identification and measurement issues but few studies have been conducted, most likely because they require the involvement of the tax authority, whose interest for increasing tax revenues might conflict with the scientific requirement of a correct randomization design. Some notable contributions among field experiment studies include Slemrod et al. (2001), Kleven et al. (2011) and Fellner-Röhling et al. (2013). In particular, Kleven et al. (2011) showed that tax evasion is small in Denmark because the share of third-party reported income is low (around 5%) and evasion rate for self-reported item is large (40%), regardless of its level of tax morality (Luttmer and Singhal, 2014). Moreover, even field experiments are not flawless because they might have limited external validity and still rely on some identification assumptions.

As convincingly suggested by Slemrod and Weber (2012), although direct measure of tax noncompliance is rare, when measuring tax evasion with a microeconomic perspective, creativity should prevail upon discouragement.

This paper looks for traces of true income by applying a well-known expenditure-based method that was developed three decades ago by Pissarides and Weber (1989) for groups of taxpayers facing different opportunities to underreport their income. Novel data was made possible by increasing the availability of admin-

[1] Estimates refer to the year 2018 for the United Kingdom (HMRC, 2019), to the year 2014 for Canada (CRA, 2016), to the year 2013 for the United States (IRS, 2019), and to the year 2017 for Italy (MEF, 2019b).

istrative data. This methodology relies on two main assumptions. First, the self-employed can underreport income and overreport tax expenditures reducing taxable income, whereas taxpayers earning third party-reported income (including mostly wage and salary workers but also contributory and non-contributory transfer receivers) have limited opportunities to underreport. Second, consumption variables – such as food expenditures – are correctly reported in surveys and they are assumed to be independent of the type of households (either self-employed or third-party-reported income receivers) after controlling for all observable characteristics. This approach has been progressively applied to several other countries (Tedds, 2008; Hurst et al., 2014; Engström and Hagen, 2017; Cabral et al., 2019; Nygård et al., 2018; Kukk et al., 2019)). However, a complete application of this method has been missing in Italy, due to the lack of reliable micro data.[2]

The application of the expenditure-based method demands the availability of detailed micro data that combine information on income and food expenditure variables. Food expenditure is often available in standard Household Budget Surveys (HBS), whose methodology is common across countries and is based on registering on a diary for a couple of weeks all expenditures by all member of each sampled household. Food expenditure is the ideal item for applying this methodology because it is always positive, with limited transitory components as opposed to other type of expenditures, is likely to be the least correlated with the self-employment status of a taxpayer, and is also likely to be well reported.[3] Typically, permanent income is proxied by current income because most household budget surveys lack the longitudinal dimension and at most provide self-declared current income, with few exceptions.[4] However, given that current income in HBSs is self-declared, it is likely to be affected by recall and measurement errors; especially when it is received irregularly, which happens to self-employed workers. It is also likely to be affected by transitory income fluctuations, due (for instance) to temporary unemployment or unexpected positive events. To reduce the attenuation bias due to measurement error of permanent

[2]In contrast to other countries, in Italy the Household Budget Survey that is conducted by the National Statistical Institute (Istat) has never included a measure of household income, even though the collection methodology and details of consumption expenditures are similar to analogous surveys conducted in other countries. The only alternative dataset, the Survey on Household Income and Wealth that is conducted by the Bank of Italy, has insufficiently detailed information in the food consumption variable asking a recall question about the average monthly spending on food in the year before the interview. Hence, the application of the expenditure-based method for underreporting measurement has been unable to provide a statistically significant result. The only available methodology for a micro-economic estimation of tax evasion in Italy was based on the "discrepancy method," which means that data from an income survey are compared to the income reported by income tax returns, assuming that taxpayers may conceal part of their income from the tax authorities but might consider declaring a higher figure to an anonymous interviewer (Bernasconi and Marenzi, 1997; Fiorio and D'Amuri, 2005; Albarea et al., 2019).

[3]Other expenditures have been used when food expenditure is not available, such as home utilities (Kukk et al., 2019; Albarea et al., 2019) or reported charitable contributions (Feldman and Slemrod, 2007).

[4]Among the few exceptions are the PSID for the United States (Hurst et al., 2014), the KLIPS for Korea Chung et al. (2010) or the RLMS for Russia Stillman and Thomas (2008).

income, a common strategy in previous studies has been to use instruments such as the education of the household head (e.g. Hurst et al., 2014; Kukk et al., 2019, among others), household characteristics (Kim et al., 2017; Gibson et al., 2008), capital income (Engström and Hagen, 2017); however, the fulfillment of their exclusion restrictions is debatable.

The recent availability of administrative data has improved the researcher's toolkit. Administrative data are longitudinal for their own nature. Therefore, average yearly income over a long period of time could be regarded as a better measure of permanent income, as long as it reduces the importance of transitory components. These measures complement HSBs when they do not provide information on current income, as is the case of Italy, which allows the estimation of an expenditure function in the search for traces of hidden income. Administrative data can also provide information on property and financial wealth. This allows a new set of possible instrumental variables to overcome the endogeneity of the income variable in the consumption equation, with arguably better chances to satisfy the exclusion restriction.

In this paper, we combine data of the HBS conducted in 2013 by the National Statistical Institute (Istat) linked with administrative records of individual tax forms, including the cadastral value of real and building property wealth, and the stock of financial wealth (bonds, shares and current accounts) for each member of the household in HBS. This link was made possible by the infrastructure of the Ministry of Economy and Finance (MEF)[5]. This enables us to provide an estimate of the traces of taxable income for the first time using the expenditure-based approach for Italy, which is a developed country with high levels of income tax evasion and a large share of self-employed workers.[6]

Our objectives in this work are twofold. In the first part, we provide the first estimation of tax evasion in Italy using the expenditure-based methodology using as measure of permanent income the average of yearly taxable income over a period of 7 years, three of which before and three after the HBS collection year. This reduces the importance of transitory components of income. This approach has been often used in the literature to build a better measure of permanent income (Solon, 1992; Gottschalk and Moffitt, 1994; Hurst et al., 2014). In this study we use average administrative data, which is free from recall and other measurement errors that are frequent in surveys using self-reported measures of income. This measure of permanent income allows us to show that the measurement errors of current income lead to an underestimation bias of the elasticity of food-expenditure, which confirms similar results by Engström and Hagen (2017). The average of the stock of individual financial wealth over a five-year period provides us with an instrument to help us get rid of remaining transitory components in the 7-year average measure of income, under the assumption that average wealth is a relevant predictor of average income but is unlikely to be correlated with the error term in the consumption equation

[5]To protect the identity of individuals and preserve anonimity, the MEF technological infrastructure is fully responsible for the exact matching of individuals in the HBS with administrative data and the data can only be analysed at the MEF's premises.

[6]`https://ec.europa.eu/eurostat/web/products-eurostat-news/-/DDN-20170906-1`.

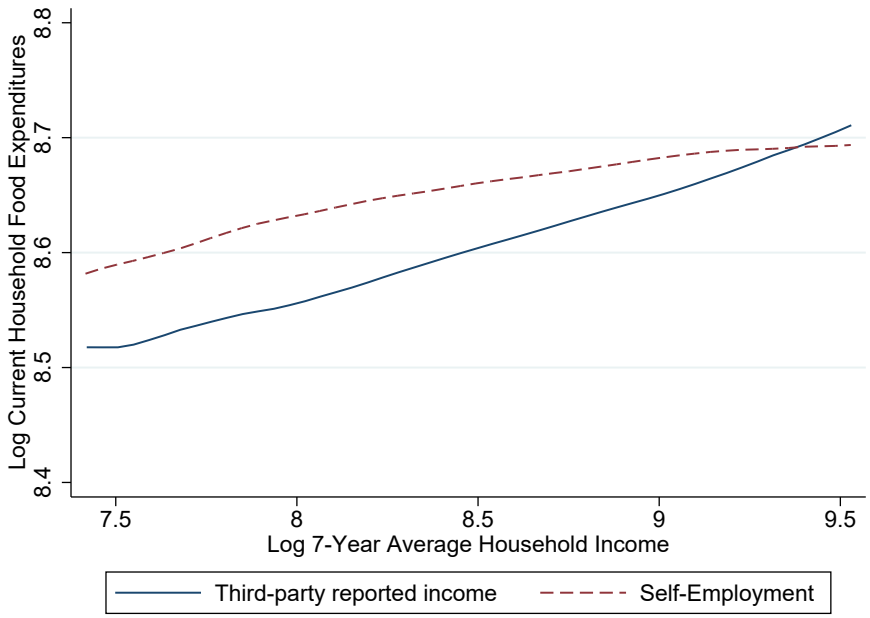
because saving is not likely to be strongly correlated with food expenditure. By exactly matching consumption and income records, including information on property and financial wealth, we got rid of statistical matching and self-reported errors, reduce measurement income to the minimum, and we manage to introduce some heterogeneity into the estimation of underreported income.

As a preliminary piece of evidence, we regressed the log of food expenditure on the log of 7-year average total household income on a vector of controls separately for each group of households (i.e., self-employed vs third-party-reporting income earners). Figure 1 plots the estimated relationship for each group derived using a nonparametric local-mean smoothing procedure. This figure shows that the food-expenditure curve estimated for self-employed consistently lies above that for third-party-reported income earners for almost all of the income support, which is consistent with the underreporting of income by self-employed households. Moreover, it also shows that the food-expenditure curve of wage and salary households is roughly constant, whereas that of self-employed households is not. This figure is similar to the findings of Hurst et al. (2014) in the United States and it overall supports the validity of the log-linear approximation of the food-expenditure curve when income is correctly measured. As far as we know, this is one of the first papers to investigate the heterogeneity in the underreporting behavior of the self-employed.[7]

Our main findings suggest that the share of unreported income among the self-employed ranges from 36-37% when the 7-year average income is instrumented using the 5-year average of financial wealth stock to 43-44% when OLS is used with the 7-year average income. These results are larger than those found with similar methodology in other developed countries; however, they rely on the definition of self-employed households. We classified self-employed households as those with at least 50% of their total income coming from self-reported income, which is closer to what would result from using self-declared occupation by the household head. However, we showed that if a lower percentage was used (i.e., the 25% that is standard in the literature), then self-employment evasion rates would be about 30%. This result is similar to what is found in the United States and is only slightly higher that what is found in the United Kingdom or even in Scandinavian countries. Furthermore, heterogeneity analysis allows us to show that households with a married head underreport less than single-headed households. However, college-educated households evade a significant share of their income regardless of their marital status, possibly because their skills and social networks provide them with better tools for cheating the government and reducing the risk of being caught.

[7] We are only aware of Cabral et al. (2019) using UK survey data and Cabral et al. (2020) using a combination of HBS and administrative data from New Zealand, although the latter mostly aims to investigate the measurement error and attenuation bias.

Figure 1: Nonparametric estimates of food-expenditures curves, by household type



This figure shows the estimates of the residuals of separate food-expenditure curves for employee and for self-employed households, conditioned on a set of observable characteristics. This figure was created by dividing the support of log income residuals into equally spaced bins and then fitting a local-mean smoothing procedure for log average income and log average expenditure for each bin. Our measure of income is the seven-year average of taxable household income. Our measure of expenditure is the current household food expenditure. We recentered the residuals at the unconditional log average income and log average expenditure for each group. Finally, we plot the data by restricting the range of the log average income residuals to be within the overlapping range for both groups. Within the overlapping range, we truncate observations with log average income below the 1st percentile and above the 99th percentile.

In the second part of this paper, we focus on the budgetary and distributional effects of self-employment underreporting by quantifying the consequences of tax evasion on public finances and inequality. As recently pointed out by Slemrod (2019, p. 4), "the real policy question is whether curbing evasion would improve the equity and efficiency implications of the public finances." The richness of our data allows for a deep investigation of the distributional effects of tax evasion, which is of growing concern (e.g. see, Nygård et al., 2018; Alstadsæter et al., 2019). The analysis is conducted by using the Italian module of EUROMOD (Ceriani et al., 2020), which is a microsimulation tax-benefit model that allows us to assess whether and to what extent the presence of tax noncompliance by self-employed in Italy produces distortions in the distribution of the tax burden.

Our results suggest that the revenue losses associated with tax evasion by

the self-employed in 2018 correspond to around 15% of total personal income tax revenues and nearly 40% of self-employed social security contributions revenues. Interestingly, our estimates of the tax gap are similar to the official estimates published by the Ministry of Economy and Finance, and were obtained by adopting a top-down approach (i.e., with a macroeconomic approach using aggregate statistics). Moreover, we find that income tax evasion hampers vertical equity, disproportionally favoring the richest taxpayers. It also undermines the redistributive impact of the tax-benefit system. Were tax evasion eradicated, then the overall disposable income inequality, as measured using the Gini index, would be about 2.5 percentage points lower than its true level (i.e., down by 6%). Our results also confirm one of the main conclusions of Hurst et al. (2014), which suggests that underreporting of income by some workers produces bias estimations of common descriptives statistics on income distribution.

This paper proceeds as follows. Section 2 discusses the methodology and presents how we capture heterogeneity in the tax evasion behavior of the self-employed relative to wage and salary earners. Section 3 describes the main characteristics of the novel data that we have used in this paper. Section 4 provides some descriptive statistics. Section 5 presents the empirical estimates of the main specification and Subsection 5.2 discusses estimated tax evasion heterogeneity. Section 6 discusses the budgetary and redistributive effects of tax evasion obtained adjusting for the estimated tax evasion. Section 7 concludes by summarizing our policy relevant findings. In the Appendix, contained in Section 8, we provide some robustness checks of our main results.

2 The methodology

Our starting point is a food expenditure equation, likewise in most of the previous works (e.g., Pissarides and Weber, 1989; Hurst et al., 2014; Kukk et al., 2019)), which describes the relationship between log-income and log-consumption for a generic household i, namely:

$$lnc_i = \alpha + \beta lny_i^P + X_i'\theta + \epsilon_i \tag{1}$$

meaning that log-household consumption, lnc_i, is proportional to *true* log-permanent income, lny_i^P, conditional on a set of observable controls, X_i. Hence, β is the income-elasticity of food demand and ϵ_i is a composite error term with mean zero. We temporarily assume that permanent income, y_i^P, is well measured in the data (the superscript P will be removed from now on for expositional convenience); we will return to this at the end of this section.

We assume different reporting behavior by households of type $j = W, S$, where W denotes third-party-income receiving households, most of which is in the form of wages and salaries but some also in the form of pensions and other transfers, and S denotes self-employed households. In particular, we assume that reported income for type-W households $\left(y_{iW}^R\right)$ is equal to their true income (y_{iW}), whereas type-S households underreport a positive amount of income $\left(y_{iS}^U\right)$, hence reporting a level lower than the true one $\left(y_{iS}^R < y_{iS}\right)$:

$$y_{iW} = y_{iW}^R \tag{2}$$

$$y_{iS} = y_{iS}^R + y_{iS}^U \tag{3}$$

Type-W households are assumed to report their income almost truthfully because their income is third-party reported and underreporting is negligible. Type-S households are allowed to self-declare their taxable income and might take the opportunity to report to tax authorities only a fraction of it.

Assuming that the share of the true income that is unreported is s for all type-S households, where $0 \leq s \leq 1$, we re-write (3) as:

$$y_{iS}^R = (1-s)y_{iS} \tag{4}$$

which allows us to rewrite (1) as:

$$lnc_{ij} = \alpha + \beta lny_{ij}^R + I_j \underbrace{[-\beta ln(1-s)]}_{\gamma} + X_i'\theta + \epsilon_{ij} \tag{5}$$

where I_j is a dichotomous variable equal to one if $j = S$ and is zero otherwise, and $\gamma = lnc_{iS} - lnc_{iW}$.

Note that from (4), using the properties of the logs and (5), one can write:

$$\kappa := \left(\frac{y_S^R}{y_S}\right) = exp\left(-\frac{\gamma}{\beta}\right), \tag{6}$$

which is the fraction of (true) income that is reported by S, as defined in Hurst et al. (2014), whose estimate is obtained as $\widehat{\kappa} = exp(-\widehat{\gamma}/\widehat{\beta})$.

Assume now that (4) also depends on other observed characteristics, Z_i, which can be a subsample of X_i, and on a set of unobserved characteristics, η_{iS}; namely,

$$lny_{iS}^R = ln(1-s) + lny_{iS} + Z_i'\delta + \eta_{iS}. \tag{7}$$

These other observable characteristics partition the dataset into ℓ mutually exclusive groups (e.g., male and female headed households, residents in the three main geographical areas of the country, etc.) and (5) becomes:

$$lnc_{ij} = \alpha + \beta lny_{ij}^R + \underbrace{I_j[-\beta ln(1-s) - Z_i'\delta]}_{\gamma_\ell} + X_i'\theta + \underbrace{\epsilon_{ij} - I_j\eta_{iS}}_{\xi_{ij}} \tag{8}$$

and (6):

$$\kappa_\ell := \left(\frac{y_S^R}{y_S}\right)_\ell = exp\left(-\frac{\gamma_\ell}{\beta}\right) \tag{9}$$

The estimation of κ_ℓ allows us to estimate the possible heterogeneity of underreporting behavior by possibly pointing at single groups showing larger share of

hidden income for further analysis and at targeting particular policies or audit activities.

In the second part of this paper, we investigate the issue of the distributional effects of hiding taxable income. In fact, once the share of the true income that is unreported, $(1-\widehat{\kappa})$, is estimated, we need to account for its relevance in terms of the share of reported self-employment income over the true total income. Some third-party reported income, y_{iWT_S}, might also be earned by type-S households, where now the subscripts i, W, T_S stand for household i, earning third-party-reported income, W, belonging to the type-S household group, respectively. We need to find by which factor we should inflate reported self-employment income by household of type-S, $y^R_{iST_S}$ without affecting their correctly reported y_{iWT_S}. For type-W households who might earn some self-employment income, y_{iST_W}, our assumptions imply that we assume that they do not conceal taxable income or if they do its amount is negligible.[8]

Ignoring from now onwards the household indicator for the sake of simplicity, let $h = y^R_{ST_S}/y^R_{T_S} = y^R_{ST_S}/(y^R_{ST_S}+y_{WT_S})$ be the share of self-employment income over total reported income. By extending what was written in eq. (4), we write:

$$y_{T_S} = \frac{1}{\kappa h} y^R_{ST_S} \tag{10}$$

$$y_{ST_S} = \frac{1-k+kh}{kh} y^R_{ST_S} \tag{11}$$

$$y^R_{T_S} = \frac{1}{h} y^R_{ST_S} \tag{12}$$

$$y^U_{ST_S} = \frac{1-\kappa}{\kappa h} y^R_{ST_S} \tag{13}$$

and consequently define four alternative tax evasion rates of type-S households, namely (i) evasion over evasion-adjusted taxable income:

$$\frac{y^U_{ST_S}}{y_{T_S}} = (1-\kappa) \tag{14}$$

(ii) evasion over evasion-adjusted self-employment taxable income:

$$\frac{y^U_{ST_S}}{y_{ST_S}} = \frac{1-\kappa}{1-\kappa+\kappa h} \tag{15}$$

(iii) evasion over reported taxable income:

$$\frac{y^U_{ST_S}}{y^R_{T_S}} = \frac{1-\kappa}{\kappa} \tag{16}$$

and (iv) evasion over self-employment reported income:

$$\frac{y^U_{ST_S}}{y^R_{ST_S}} = \frac{(1-\kappa)}{\kappa h} \tag{17}$$

[8] Indeed, the median share of self-employment income for type-W households in our sample is zero and the average share is 1.5%.

Equation (17) highlights that as the share of self-employment income in the economy increases ($h \to 1$), the share of evasion over total reported self-employment income increases. This suggests the need to introduce different but interacting policies to curb tax evasion. Increasing κ implies reducing the opportunities of type-S households to hide part of their income, increasing h implies reducing the opportunity to complement self-reported with third-party-reported income.

Let us now return to the issue of the correct measure of permanent income. The correct identification of κ or κ_ℓ relies on the assumption that we introduced in (1) that households want to smooth out food consumption over time, as an effect of the diminishing marginal utility assumption. This means that household food expenditure in a given point in time is not determined by its current income but is a consequence of consumption smoothing, where households spread out windful gains or unexpected losses over time and transitory changes in income are expected to have only small effects. This means that current income as a measure of permanent income would be affected by transitory shocks, biasing the β coefficient estimate towards zero. Attenuation bias is likely to be an even larger issue if information on current income was self-declared in a survey interview and not obtained from administrative data, increasing the measurement error due to recall or mis-measurement problems. Averaging current income over a long period of time can better capture the permanent component of income using current income (Solon, 1992; Gottschalk and Moffitt, 1994); however, this is often unfeasible for most available surveys (i.e., standard HBS such as the CE in the United States or national HBS in EU-member states). However, even the several-year average current income would not be a perfect measure of permanent income because some transitory component might still be present.

Instrumental variable (IV) methods are often used to mitigate the attenuation bias. Given that education is a key determinant of market income, the education of the household head is often used as an IV (e.g. Hurst et al., 2014; Kukk et al., 2019) assuming that educational attainments only affect consumption by affecting permanent income. This is, however, a strong assumption because there exists a large body of evidence suggesting that education is a key determinant of household consumption, and in particular of food-expenditure (e.g. see Ricciuto et al., 2006; Turrell et al., 2003; Worsley et al., 2004). Engström and Hagen (2017) recently used capital income under the assumption that food-expenditure is mostly a function of labor income. Here, we will make use of stock of wealth, claiming that it satisfies the orthogonality assumption in the consumption equation because households typically do not exchange their assets for food. However, the problem is often a data availability issue: information on individual financial wealth is often limited for privacy reasons, and real estate and building property wealth is often mis-measured because administrative data often report cadastral values only, which can greatly differ from actual market values.

The administrative data that has recently been made available provide new and important tools in the researcher's kit. Administrative data measure income through tax forms without recall errors. In administrative data, one observes

only reported income, in contrast in many surveys the respondents are allowed to declare something different from what was presented to the tax authority; hence, they could possibly include disposable income that was partly evaded, which makes measurement error larger and its bias larger.[9]. As discussed in Section 3, in this paper we make use of a unique dataset combining the 2013 wave of the Italian HBS, with individual tax data between 2010 and 2016. This allows us to use the 7-year average of current incomes as a measure of permanent income and contrast it with the 2013 current income by assessing the size of part of its transitory income component. Furthermore, by linking an individual household member to their individual stock of financial wealth (bonds and shares), we can instrument the 7-year average income by cleansing the average income for remaining transitory income components.

3 The data

To find traces of true income, we built a novel consumption-income dataset by linking a representative sample of Italian households' consumption patterns (the 2013 Italian HBS) with administrative data. The link was performed at the individual level for all members of surveyed households by the Italian Ministry of Economy and Finance (MEF)'s technological infrastructure to respect the full anonymity of taxpayers who responded to the survey.

The Household Budget Survey

The Italian HBS is produced by the Italian National Institute of Statistics and it is run yearly to provide a picture of the consumption patterns of Italian households. The HBS is used for several purposes at the national level, including the construction of Italian National Accounts, the measurement of absolute and relative poverty at national level, and the measurement of inflation by household expenditure classes (Rondinelli, 2014). The Italian HBS contains information on 278 elementary consumption and self-consumption items, imputed rent, and data on demographic and economic household characteristics. The main expenditure variable that we use as a dependent variable in the empirical analysis is the monetary value (in euros) of total food consumption expenditures that are recorded in the HBS on a daily basis from a diary kept by a member of the household for two weeks. This expenditure category includes information on 55 food expenditures consumed by any household member at and away from home, including self-produced and auto-consumed goods, and expenses at bars, restaurants and canteens. In the empirical analysis, as control variables, we use all of the key demographic and economic information available in the HBS, such as household composition (e.g., single adults vs couples, number of children),

[9] In fact, the discrepancy between income reported in an anonymous survey and in tax records provides valuable information that has been used to find traces of reported income, assuming that respondents to the survey report a closer-to-true income than to tax authorities (Bernasconi and Marenzi, 1997; Fiorio and D'Amuri, 2005; Kukk et al., 2019) following the so-called "discrepancy method" Nygård et al. (2018)

the educational level of the head and of the partner (if present), age, and the gender of the household head.

Using the personal identification tax code, HBS individual records are linked to the administrative archives of individual tax forms and individual financial wealth stocks.[10] Households with missing observations were dropped, causing a loss of about 12% of all HBS households and providing us with a sample of 44,284 individuals living in 18,198 households.[11]

Tax record administrative data

We measure current before- and after-tax income using individual tax records, which allows us to focus on the actual reported income. The panel structure of tax data allows us to construct a measure of reported individual income by averaging current income from year $t-3$ to year $t+3$, where $t = 2013$. This allows us to reduce the relevance of the transitory component and improve in terms of permanent income approximation.

Tax forms also provide information on the province of residence of taxpayers. This allows us to cluster standard errors and to account for any possible correlation within the 109 Italian provinces. Tax record data also provide information on real estate and building property wealth at the individual level. However, because it is recorded at cadastral values (which are only a noisy measure of the actual market values because most of them were defined in the 1960s and have never been updated since), in the empirical analysis we only add the 7-year average cadastral value of properties as a control variable (over the period 2010-2016) to control for part of the unobserved heterogeneity.

Financial accounts administrative data

This combined dataset is further enriched by an additional administrative data source that collects information on financial wealth. Since 2012, in Italy, all bank and financial institutions must provide information on individual financial accounts (e.g., bank deposits, credit cards, investment funds, portfolios, derivatives, etc.) to the Italian tax administration on an annual basis. Since 2019, these data can be used by the MEF for conducting analyses on tax evasion and tax revenues, while preserving individual anonimity.[12] We use the average total financial wealth for each taxpayer by aggregating multiple individual financial accounts owned by the same taxpayer [13] over the 2014-2018 period, dropping

[10] Although similar link was performed by Engström and Hagen (2017) using Swedish data, in their setting full administrative data are only available for household heads and a negligible share of the other members of the household responding to the HBS.

[11] Most of the missing values were due to typos in the compilation of the identification tax code in the survey and to tax form exemption for a minority of taxpayers, mostly regarding pensioners with no other source of income. We assume that these missing values are uncorrelated with the income type (self- or third-party reported) earned.

[12] For more information, see `https://www.agenziaentrate.gov.it/portale/web/guest/schede/comunicazioni/integrativa-archivio-dei-rapporti-con-operatori-finanziari`.

[13] The average taxpayer has shares in more than three financial accounts.

the first two years of data collection because of data quality concerns. Once again, we aim to reduce the role of temporary fluctuations by taking into account an average value of the household financial wealth, which is made possible by the panel structure of the administrative data.

4 Descriptive statistics

Given that the HBS collects consumption at the household level, the consumption equation is estimated at the household level, which is also the dimension that is commonly used in distributional analyses. Similarly, we present the descriptive statistics of the main variables used in the empirical analysis.

We classified a household as self-employed if 50% of its income comes from self-employment. The aim of using a proportion of reported income is common in the literature and, since Pissarides and Weber (1989), a share of 25% is more often used. Although most households earn either self-employed or third-party reported income (wage and salary, contributory and contributory transfers), about 15% of our sample earn both self-employed and third-party reported income. A classification of self-employed households using a larger share will allow us to be more conservative, which reduces the likelihood to classify some of them as self-employed when they are in fact mostly third-party reporting households. Moreover, using the 50% share as a main threshold allows us to count a number of self-employed households close to that emerging from using the self-declared occupation of the household head in the HBS. Although we believe that using the 50% share to identify self-employed households is better suited to the Italian context, we will briefly present a broad range of some robustness checks based on different criteria to classify households as self-employed. We will also include the self-declaration of the household head, or age and gender of the household head.

We do not disentangle between self-employed workers and unincorporated businesses to distinguish between "entrepreneurs" and other business owners. Although they might be characterized by distinct cognitive and non-cognitive traits (Levine and Rubinstein, 2017), they share a similar opportunity to evade because they are both allowed to largely self-report their income.

In our analysis, we do not select households on any dimension—namely, we opted to keep all households in the estimation sample and use control variables to correct for heterogeneity preserving the representativeness of the whole sample to the total population—because one of the purposes of our analysis is to assess the effect of underreported income on the whole population of households and not on a selected part of the population.

Tables 1 and 2 show the mean value and its standard error for the main variables used in the empirical analysis for the whole sample, group of self-employed households and the remaining households, earning third-party reported income. Income data are presented as pre-tax and post-tax total household income, which aims to highlight the smoothing effect of the Italian progressive tax and benefit system for the estimation of tax evasion.

In Panel (A), we report the average level of log food and log home utilities expenditures and their standard errors. Since Pissarides and Weber (1989), food expenditure has been used as an ideal dependent variable candidate because it is assumed to be uncorrelated with the self-employment status of a household, holding constant all other observable characteristics. In recent papers (e.g. Paulus (2015), Albarea et al. (2019)) home utility expenditures were used instead of food expenditures because the former are usually available in surveys on household income, including the EU-SILC. However, in our view, using home utilities expenditure as dependent variable in eq. (1) is not ideal because home utilities expenditures are likely to be correlated to the self-employment status of the households because some self-employed workers might, to some extent, work from home or claim utilities as business expenditures, which would affect their total amount in a way that the data could not clarify. However, because the HBS also has detailed information on home utility expenditures, we will use food expenditure as our main consumption item and utility expenditure as an alternative dependent variable to test the robustness of our main results. These robustness checks will be briefly presented in the Appendix.

The descriptive statistics in Panel (A) show that the average level of food and home utilities consumption for self-employed households are, in absolute levels, roughly 18% and 11% larger than for third-party reported income households respectively. However, Panel (B) shows that self-employed households declare on average about 12% less total current income than type-W households, regardless of income being measured as pre-tax or as post-tax, which reduces to 7% less when 7-year average income is used instead. Standard errors tend to be larger for self-employed households, although this is partly a consequence of a significantly smaller sample size as opposed to the third-party-reported income group. Panel (C), in Table 2, shows other characteristics that are used as controls, regarding the household, the household head and education of the head's spouse, some heterogeneity can be seen between self-employed and wage and salary households. In particular, the consumption of some sin goods, which could be seen as proxies of propensity to risk, tend to be either non-significantly different or larger (e.g., tobacco products and spirits) for self-employed households. Female are less likely to be the head of self-employed households, who seem to have (on average) larger, younger and more educated heads.

Table 1: Descriptive statistics, main variables.

	Whole sample	**Self-employed**	**Third-party reported income**
	A. Log average expenditure measures		
log food expenditures	8.622	8.768	8.606
	(0.004)	(0.014)	(0.005)
log home utilities expenditures	7.732	7.827	7.722
	(0.004)	(0.014)	(0.005)
	B. Log average income measures		
Pre-tax Total Household Income:			
current (in logs)	10.088	9.974	10.101
	(0.007)	(0.030)	(0.007)
average (7-year, in logs)	10.098	10.037	10.105
	(0.006)	(0.026)	(0.006)
Post-tax Total Household Income:			
current (in logs)	9.914	9.798	9.927
	(0.006)	(0.028)	(0.006)
average (7-year, in logs)	9.920	9.848	9.929
	(0.006)	(0.024)	(0.006)
Sample size	18198	1767	16431

Notes: our calculation on the selected sample, standard errors in parentheses. Self-employed households are identified as those with self-employment income equal or larger to 50% of total household income.

Table 2: Descriptive statistics, other household control variables.

	Whole sample	**Self-employed**	**Third-party reported income**
	C. Other Household measures		
% of female headed Households	0.320	0.191	0.335
	(0.003)	(0.009)	(0.004)
% families with kids	0.264	0.307	0.259
	(0.003)	(0.011)	(0.003)
Average Household size	2.377	2.852	2.323
	(0.009)	(0.032)	(0.010)
Household head: 35 and below	0.075	0.096	0.073
	(0.002)	(0.007)	(0.002)
Household head: 36-50	0.290	0.499	0.266
	(0.003)	(0.012)	(0.003)
Household head: 51-65	0.281	0.333	0.276
	(0.003)	(0.011)	(0.003)
Household head: 66 and over	0.353	0.073	0.385
	(0.004)	(0.006)	(0.004)
No Couple	0.434	0.332	0.446
	(0.004)	(0.011)	(0.004)
In-couple × Primary Educ. Spouse	0.283	0.258	0.285
	(0.003)	(0.010)	(0.004)
In-couple × Secondary Educ. Spouse	0.212	0.311	0.201
	(0.003)	(0.011)	(0.003)
In-couple × College Educ. Spouse	0.071	0.099	0.068
	(0.002)	(0.007)	(0.002)
North	0.498	0.542	0.493
	(0.004)	(0.012)	(0.004)
Center	0.205	0.187	0.207
	(0.003)	(0.009)	(0.003)
South	0.297	0.272	0.300
	(0.003)	(0.011)	(0.004)
Sin goods: Tobacco Products	1.726	2.161	1.676
	(0.022)	(0.075)	(0.023)
Sin goods: Lottery Tickets	0.674	0.678	0.673
	(0.014)	(0.044)	(0.014)
Sin goods: Spirits	0.573	0.607	0.569
	(0.013)	(0.043)	(0.014)
Primary Educ. Household Head	0.534	0.395	0.550
	(0.004)	(0.012)	(0.004)
Secondary Educ. Household Head	0.334	0.406	0.326
	(0.003)	(0.012)	(0.004)
College Educ. Household Head	0.131	0.199	0.124
	(0.003)	(0.010)	(0.003)
Property wealth (cadastral values, in logs)	4.891	5.339	4.840
	(0.021)	(0.065)	(0.022)
Financial wealth (in logs)	8.845	8.737	8.858
	(0.023)	(0.087)	(0.024)

Notes: as in Table 1.

5 The estimates

In this section we present results of the average share of unreported income by type-S households (Subection 5.1), using eq. (5), and of some mutually exclusive groups to detect underreporting heterogeneity (Subsection 5.2), using eq. (9).

5.1 The average underreporting

In Table 3, we provide estimate of eq. (5) for both pre-tax and post-tax total family income and using as measure of permanent income, $ln\, y_i$, either the current income or the average over the 7-year period. Columns (A)-(D) differ for the use of different set of controls. A first striking result of Table 3 is that the high level of precision of estimates suggest that by removing possible recall and measurement errors of standard income surveys by linking administrative income data to HBS, self-employed households consume on average consistently more food than wage and salary households, thus showing traces of unreported income. The first two columns use no controls, namely log consumption is regressed on a constant, the self-employment dummy and the log of income. In the following two columns, we include as controls gender and age of the household head, in-couple dummy interacted with education (primary, secondary or college) of the partner, household size, a dummy for presence of kids, dummies for main area of residence (North, Center or South), total consumption of sin goods, building property wealth (measured using the 7-year average total cadastral values in logs) and education of the household head. The last two columns show the instrumental variable estimation using the stock of financial wealth (bonds and shares) as an instrument. All of the standard error estimates are adjusted for 109 clusters at the province level of residence of the family.

The use of the 7-year average income as opposed to the current income as measure of permanent income has no significant effect on the self-employment dummy, which suggests that the income transitory component has no effect on the choice of being self-employed. In contrast, the effect is significant on the estimation of the income-elasticity of food demand, which suggests that there is a larger transitory income components present in current income bias the estimation towards zero, as expected. The estimates suggest that type-S households consume on average 5-6% more than type-W (columns C and D).

By using equation (5) and estimating the proportion of unreported income as $(1-\widehat{\kappa})$, we conclude that self-employed households in Italy underreport on average over 40% of both their pre-tax or post-tax total family income relative to third-party-reported income receivers. These results are consistent with the official estimates for Italy published by MEF (2019b), and when looking at underreported income by (Braiotta et al., 2020), with the use of national accounts and the top-down measurement approaches (for a review of the top-down approaches, see HMRC, 2019).

Columns (E) and (F) report estimates using (log) stock of financial wealth as an instrument for the measure of permanent income used, current and 7-year

average respectively. The first stage F-statistic is large, which confirms that the instrument is relevant. As expected, our results are now close, regardless of the measure of income being instrumented. More importantly, they suggest that by cleansing the temporary component of income using the stock of wealth as instrument, the income-elasticity of food demand is above 15% for pre-tax income and about 17% for post-tax income, which is more than 50% larger than OLS estimates. The results using IV also show that the estimates of the self-employment dummy are larger than in the OLS case, reaching an estimate of $(1-\kappa)$ slightly lower, just below 40%.

Although this share might seem large as opposed to what was found in most other developed countries, it relies on our definition of self-employed households (i.e., those earning 50% or more from self-employment). To test the robustness of this conclusion, we present Table 4, which provides estimates changing the classification of type-S households but using the same sample as was used previously, presenting only results using the 7-year average measure of permanent income and using OLS and IV with full controls only.[14] Model 1 shows that had the share been fixed at the 25%, as in most previous studies, then the share of hidden income would reduce to around 30%, which is in line with that found in the United States (Hurst et al., 2014), and is only slightly higher than that found (for instance) in the United Kingdom (Cabral et al., 2019) or in Sweden (Engström and Hagen, 2017). Models 2 and 3, which rely on the classification of a type-S household as one that earns any or all its income from self-employment, respectively, suggest that the classification of self-employed households can provide very different results in terms of estimation of traces of taxable income. Focusing on IV results, according to Model 2 type-S households are more than 20% and on average their evasion rate is about 23%; while according to Model 3, there is a very small share of type-S households but they evade around 80% of their income. Finally Model 4, which uses the self-reported classification of the household head to define a household as of type-S or of type-W, suggests that the results using self-reported classification (as declared in the HBS) are closer to our preferred classification both in terms of type-S population share and in terms of estimated tax evasion, $(1-\widehat{\kappa})$. Additional robustness checks regarding the classification of type-S households and introducing some additional sample selection are presented in the Table 11 in the Appendix.

[14] In other words, we show in Table 4 the estimation of four models which are equal to those presented in Table 3, columns D and F except for the definition of the dependent variable.

Table 3: Food (log) expenditure equation with alternative measures of permanent income.

	(A) OLS	(B) OLS	(C) OLS	(D) OLS	(E) IV	(F) IV
	Income = Pre-tax Total Family Income					
Self-employed	0.187***	0.177***	0.053***	0.055***	0.077***	0.071***
	(0.017)	(0.017)	(0.016)	(0.017)	(0.024)	(0.022)
Current inc.	0.197***		0.076***		0.157***	
	(0.009)		(0.008)		(0.051)	
Aver. inc. (7-year)		0.233***		0.094***		0.154***
		(0.009)		(0.009)		(0.048)
$(1-\kappa)$	0.612***	0.534***	0.501***	0.441***	0.389***	0.369***
	(0.038)	(0.037)	(0.107)	(0.098)	(0.070)	(0.075)
Controls	No	No	Yes	Yes	Yes	Yes
R-squared	0.097	0.115	0.261	0.263	0.251	0.259
N. obs	18198	18198	18198	18198	18198	18198
N. obs Self-Employed	1767	1767	1767	1767	1767	1767
Share Self-Employed	0.866	0.866	0.866	0.866	0.866	0.866
F-stat					176.629	207.798
	Income = Post-tax Total Family Income					
Self-employed	0.189***	0.183***	0.052***	0.056***	0.080***	0.074***
	(0.017)	(0.017)	(0.016)	(0.017)	(0.025)	(0.023)
Current inc.	0.216***		0.077***		0.173***	
	(0.011)		(0.009)		(0.056)	
Aver. inc. (7-year)		0.259***		0.099***		0.168***
		(0.010)		(0.011)		(0.053)
$(1-\kappa)$	0.584***	0.506***	0.492***	0.432***	0.372***	0.357***
	(0.038)	(0.035)	(0.107)	(0.096)	(0.066)	(0.070)
Controls	No	No	Yes	Yes	Yes	Yes
R-squared	0.098	0.119	0.26	0.262	0.249	0.257
N. obs	18198	18198	18198	18198	18198	18198
N. obs Self-Employed	1767	1767	1767	1767	1767	1767
Share Self-Employed	0.866	0.866	0.866	0.866	0.866	0.866
F-stat					169.646	205.36

Notes: A self-employed household has at least 50% of its income from self-employment. Controls include household head age and gender, in-couple dummy interacted with education (primary, secondary or tertiary) of the spouse, household size, a dummy for presence of kids, family consumption of sin goods, a full set of macro area of residence dummies, household head education and building property wealth (cadastral values). IV estimation uses (log) financial wealth as instrumental variable.
Standar errors are adjusted for 109 clusters at the province of family residence.

Table 4: Expenditure equations for (log) home utility expenditures and alternative measures of permanent income using different definitions of self-employed households over the same sample.

	Model 1		Model 2		Model 3		Model 4	
	(A)	(B)	(C)	(D)	(E)	(F)	(G)	(H)
	OLS	IV	OLS	IV	OLS	IV	OLS	IV
	Income = Pre-tax Total Family Income							
Self-employed	0.042***	0.054***	0.034***	0.041***	0.188***	0.257***	0.061***	0.074***
	(0.015)	(0.019)	(0.010)	(0.012)	(0.063)	(0.083)	(0.017)	(0.021)
Aver. inc. (7-year)	0.094***	0.153***	0.093***	0.151***	0.095***	0.157***	0.094***	0.152***
	(0.009)	(0.048)	(0.009)	(0.048)	(0.009)	(0.049)	(0.009)	(0.048)
$(1-\kappa)$	0.360***	0.300***	0.308***	0.236***	0.860***	0.805***	0.475***	0.386***
	(0.107)	(0.074)	(0.077)	(0.062)	(0.092)	(0.089)	(0.094)	(0.083)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.263	0.259	0.263	0.259	0.263	0.258	0.263	0.259
N. obs	18198	18198	18198	18198	18198	18198	18198	18198
N. obs Self-Employed	2368	2368	3928	3928	124	124	1995	1995
Share Self-Employed	0.653	0.653	0.474	0.474	1.000	1.000	0.537	0.537
F-stat		211.402		217.506		212.408		221.191
	Income = Post-tax Total Family Income							
Self-employed	0.042***	0.057***	0.035***	0.043***	0.190***	0.269***	0.061***	0.076***
	(0.015)	(0.020)	(0.010)	(0.012)	(0.064)	(0.086)	(0.018)	(0.021)
Aver. inc. (7-year)	0.098***	0.167***	0.097***	0.165***	0.100***	0.173***	0.099***	0.166***
	(0.011)	(0.053)	(0.011)	(0.052)	(0.011)	(0.053)	(0.011)	(0.052)
$(1-\kappa)$	0.351***	0.290***	0.303***	0.229***	0.850***	0.790***	0.460***	0.366***
	(0.104)	(0.069)	(0.075)	(0.057)	(0.094)	(0.088)	(0.093)	(0.078)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.262	0.257	0.262	0.257	0.262	0.257	0.262	0.258
N. obs	18198	18198	18198	18198	18198	18198	18198	18198
N. obs Self-Employed	2368	2368	3928	3928	124	124	1995	1995
Share Self-Employed	0.653	0.653	0.474	0.474	1.000	1.000	0.537	0.537
F-stat		209.137		215.859		210.827		219.629

Notes:
Model 1: A self-employed household has at least 25% income from self-employment.
Model 2: A self-employed household has any positive income from self-employment.
Model 3: A self-employed household has 100% income from self-employment.
Model 4: A self-employed household has a head who self-declares to be self-employed in the survey.
Controls include household head age and gender, in-couple dummy interacted with education (primary, secondary or tertiary) of the spouse, household size, a dummy for presence of kids, family consumption of sin goods, a full set of macro area of residence dummies, household head education and building property wealth (cadastral values).
Standar errors are adjusted for 109 clusters at the province of family residence.

5.2 Heterogeneity of income underreporting

Underreporting has typically been estimated assuming that all of the self-employed hide a constant amount of their true income, although descriptive evidence such as our Figure 1 (and Figure 1 in Hurst et al., 2014) suggests that heterogeneity could play an important role. The importance of sociodemographic characteristics in tax compliance behavior have also been discussed in Hofmann et al. (2017). The richness of our data, as opposed to available data in other countries, allows us to test heterogeneity by assuming that the fraction of the true income that is reported depends on some observable characteristics, as in eq. (9). In Table 5, we test the significance of such heterogeneity by estimating eq. (8) for both pre- and post-tax total family income, using the 7-year average measure of income as a measure of permanent income, the whole set of controls and a set of interactions of the self-employed household dummy with education of the household head, and with a dummy to indicate whether the household includes a couple or a single adult.

In the first two columns, we present estimates using pre-tax total family income and in the last two estimates using post-tax total family income. When IV is used, we observe that the IV variable (namely, the log of financial wealth) is very significant, with an F-statistic above 200. This provides an estimate of the food elasticity between 15% and 17%, which is about 50% larger that what would have been estimated using OLS.

Accounting for heterogeneity, Table 5 shows that, *ceteris paribus*, self-employed households consume about 17% more than type-W households but some heterogeneity emerges. In particular, in-couple self-employed households have a consumption pattern that is closer to that of third-party-reported income households because their average coefficient is negative. This means that their coefficient γ_ℓ in eq. (9) will be close to zero, suggesting that there is no evidence of a significant trace of income hidden to the tax authorities. There might be various reasons why in-couple self-employed households do not seem to evade taxes. For instance, in-couple self-employed households are more likely to have a larger share of third-party-reported income, hence mitigating the chances of tax evasion as part of the income would be third-party reported; in-couple household are more likely to have children and larger responsibilities and might be more risk averse; in-couple households might increase the likelihood of having access to the welfare system for income support and in-kind benefits promoting a positive attitude towards redistributive role of the State; and, in-couple households might face coordination problems, which increases the chances that one of the two partners would oppose cheating the government. However, these issues seem to be smaller issues for college educated households. In fact, college-educated households, regardless of their marital classification, show a similar propensity to conceal income as the single, primary-educated households (who are the reference category). One may argue that cheating the government by reducing payable taxes is easier if you know the best accountant or are in networks with well-qualified consultants who can advise on how to minimize the chances of being discovered. One could also argue that the highly-educated self-employed

work in service sectors (e.g., liberal professionals), where evasion opportunities are higher. However, the relevance of these alternative explanations are left for future research.

In Table 6, we compute the share of evaded pre-tax income, $(1 - \kappa_\ell)$, according to the groups defined by interactions of Table 5, similarly to Cabral et al. (2019). Focusing on IV results, it can be seen that most of the evasion of self-employed income comes from single-headed households, who evade from half (high school educated) to nearly 70% of their income (primary school and college educated). In-couple self-employed households seem to evade between about nothing to 40% among college educated headed households.

Table 5: Food (log) expenditure equation using 7-year average as measure of permanent income and estimating heterogeneity among the self-employed.

	(A)	(B)	(C)	(D)
	Income = Pre-tax Total Family Income		**Income = Post-tax Total Family Income**	
	OLS	IV	OLS	IV
Average inc. (7-year)	0.092***	0.153***	0.097***	0.167***
	(0.009)	(0.048)	(0.011)	(0.053)
Self-employed (SE)	0.164***	0.173***	0.166***	0.178***
	(0.039)	(0.040)	(0.039)	(0.041)
SE × Single × Second. educ.	-0.089*	-0.069	-0.089*	-0.068
	(0.050)	(0.054)	(0.050)	(0.055)
SE × Single × College educ.	-0.005	-0.002	-0.005	-0.002
	(0.056)	(0.058)	(0.056)	(0.058)
SE × Couple × Primary educ.	-0.119***	-0.116***	-0.120***	-0.117***
	(0.044)	(0.045)	(0.044)	(0.044)
SE × Couple × Second. educ.	-0.195***	-0.189***	-0.198***	-0.192***
	(0.046)	(0.047)	(0.046)	(0.047)
SE × Couple × College educ.	-0.088	-0.087	-0.089	-0.089
	(0.058)	(0.056)	(0.058)	(0.056)
R-squared	0.264	0.26	0.263	0.258
N. obs	18198	18198	18198	18198
N. obs Self-Employed	1767	1767	1767	1767
Share SE income	0.866	0.866	0.866	0.866
F-stat		205.019		202.661

Notes: A self-employed household has at least 50% of its income from self-employment. Controls include household head age and gender, in-couple dummy interacted with education (primary, secondary or tertiary) of the spouse, household size, a dummy for presence of kids, family consumption of sin goods, a full set of macro area of residence dummies, household head education and building property wealth (cadastral values). IV estimation uses (log) financial wealth as instrumental variable.
Standard errors are adjusted for 109 clusters at the province of family residence.

Table 6: Estimates of share of unreported income $(1-\kappa_\ell)$.

	(A)	(B)
	Income = Pre-tax Total Family Income	
$(1-k_\ell)$, where ℓ is:	OLS	IV
SE × Single × Primary educ.	0.831***	0.679***
	(0.079)	(0.124)
SE × Single × Second. educ.	0.555**	0.494***
	(0.225)	(0.155)
SE × Single × College educ.	0.822***	0.674***
	(0.092)	(0.138)
SE × Couple × Primary educ.	0.382**	0.312***
	(0.177)	(0.120)
SE × Couple × Second. educ.	-0.405	-0.105
	(0.440)	(0.251)
SE × Couple × College educ.	0.562***	0.434**
	(0.212)	(0.183)

Notes: This table displays $(1-\kappa_\ell)$ computed as in eq. (9) for model specification of Table 5. A self-employed household has at least 50% of its income from self-employment.

6 The redistributive effects of income underreporting

The shares of unreported income estimated in the first part of the paper taking into account the socio-demographic characteristics of the taxpayers (Table 6 column B) are used to simulate social contributions, taxes and benefits under different scenarios of individual tax compliance with a double aim. First, we assess the budgetary impact of tax evasion by the self-employed by quantifying the revenue losses associated to evaded taxes and social contributions. For a correct accounting of the lost revenues, including both taxes and social security contributions paid by self-employed (SSC), we apportion the traces of hidden income found to self-employment income only using equation 11; hence, we obtain an evasion-adjusted version of self-employment income for type-S households, y_{ST_S}. Second, we provide a quantification of the effects of income tax evasion by the self-employed on vertical equity, which highlights the extent to which tax evasion affects income inequality, undermines the redistributive impact of the tax-benefit system and disproportionally favors some taxpayers.

Our starting point is a scenario where only incomes declared to the tax authorities are considered and no income is concealed (Reported income scenario). In the second scenario, we assume that the self-employed underreport their incomes to the extent estimated in the first part of the paper (i.e., assuming that

on top of what is reported to the tax authorities, the self-employed can also enjoy an extra amount of disposable income by concealing it: true income scenario). In the third scenario, the self-employed are assumed to declare their full incomes (i.e., we assume that evasion, as estimated here, is fully discovered and income is adjusted so to obtain full tax compliance: evasion-adjusted scenario).

The simulation of taxes and benefits in the different scenarios is performed using EUROMOD, which is the EU-wide tax-benefit microsimulation model (Sutherland and Figari, 2013) and has been extensively validated and used in academic and policy related analyses. The policy rules considered are those in force in 2018. Income values, taken from 2013 tax data, have been uprated to 2018 by using specific updating indexes (Ceriani et al., 2020). EUROMOD is used as a static model (i.e., simulations are performed assuming no changes in labor supply behavior of taxpayers under the different scenarios of tax compliance).

Table 7 provides the set of tax evasion rates discussed in Section 2 that one could reasonably compute assuming constant or heterogeneous evasion rates among type-S households, namely using $(1-\widehat{\kappa})$ as in the last column of Tables 3 and 5, respectively. This shows that type-S households in Italy evade from about 40% of their total evasion-adjusted income, which rises to around 80% if measured relative to what they report to the tax authority, $y^R_{ST_S}$. By taking heterogeneity into account, the evasion rates are slightly larger than those derived assuming constant share of underreporting. This highlights the importance of exploiting the richness of the available data.

The first column of Table 8 shows the main aggregates of "reported income scenario," which well reflect the latest available official statistics (MEF, 2019a). This confirms that the sample used in the analysis is a reliable representation of the national population of taxpayers and the simulations capture well the tax-benefit rules in force in the country. It shows that self-employment income amounts to about €105 billion, which is about 12.5% of the total taxable income (€830 billion), and the social security contributions (SSC) paid by self-employed about to €22.5 billion. The second column of Table 8 provides aggregate values under the "true income scenario", showing that the expenditure-based methodology adopted in this paper to identify traces of concealed disposable income total about €72 billion of concealed income. As shown in the bottom panel of Table 8, this amounts to a sizable share of disposable income (11%).

Finally, the third column illustrates the "evasion-adjusted scenario" in €million (top panel) and in percentage with respect to reported income (bottom panel). Had the estimated concealed income been totally included in the self-employment tax base, then declared self-employment income would have increased by 68% and disposable income by 6%. The loss of personal income tax (Irpef) would be 16% and that of self-employment SSC to 39%.[15]

Thanks to the limited generosity of means-tested transfers (the national minimum guaranteed income was only introduced in 2019), the additional costs

[15]The proportional difference of Irpef and SSCs loss revenues is a direct consequence of the fact that SSCs are proportional, with a cap at €6,500 monthly, whereas Irpef is progressive.

of social benefits due to income underreporting to the tax authorities are almost negligible. Interestingly, when individuals are ranked according to the their equivalized-disposable income (using the OECD-adjusted equivalence scale) in the evasion-adjusted scenario, the percentage variations of disposable income are clearly concentrated at the top of the distribution (Figure 2). Had we used the reported income to rank individuals, we would also have observed important variations in the disposable income at the bottom of the distribution because evasion-adjustment produces important re-ranking and causes tax evaders to move up the income ladder (Johns and Slemrod, 2010).

The distribution of revenue losses (by decile groups of evasion-adjusted equivalized disposable income) is clearly right-skewed with around 80% of the missing tax revenue attributable to the taxpayers at the top decile of the distribution. The distribution of the revenue losses related to social security contributions is a bit flatter due to their proportional nature and the existing cap (Figure 3). The concentration of total revenue losses at the top of the distribution is a consequence of the unequal distribution of self-employment income and the overall progressivity of the tax system. The latter is responsible for the uneven gains of income concealment, with taxpayers in the first half of the distribution saving from 5 to 9 percentage points in effective tax rates and the richest taxpayers paying saving up to 23 percentage points as opposed to a scenario of full tax compliance (Table 9).

The consequences of tax evasion in terms of income inequality are not negligible (Table 10). Italy is characterized by a relatively high level of disposable income inequality and a reduced redistributive impact of the tax-benefit system. Estimates based on reported administrative data show a Gini index of around 0.56 for market income and 0.35 for disposable income, with a redistributive effect of the tax-benefit system measured by the Reynolds-Smolenski index equal to 0.21. However, considering the true income scenario, the inequality of market and disposable income would increase to 0.58 and 0.38, respectively, and redistribution would diminish to 0.19. In a scenario of full tax compliance (evasion-adjusted income), the Gini index of disposable income would be midway between the reported and the true scenario. This suggests that inequality measures based on reported income might be underestimated, as also discussed by Hurst et al. (2014).

Table 7: Tax evasion rates with and without heterogeneity of self-employed households, in percentage.

	Evasion over evasion-adjusted taxable income $y^U_{ST_S}/y_{T_S}$	Evasion over evasion-adjusted self-employment income $y^U_{ST_S}/y_{ST_S}$	Evasion over reported taxable income $y^U_{ST_S}/y^R_{T_S}$	Evasion over reported self-employment income $y^U_{ST_S}/y^R_{ST_S}$
	(A)	(B)	(C)	(D)
No heterogeneity	37.56	42.72	60.16	74.57
With heterogeneity	40.78	46.05	68.87	85.37

Notes: Tax evasion rates are presented in Section 2. Results without heterogeneity consider results of Table 3. Results with heterogeneity are built using results of Table 6.
Source: authors' elaboration based on EUROMOD (HBS-fiscal data).

Table 8: Budgetary effects of different tax compliance scenarios, 2018

	Reported (A)	True (B)	Evasion-adjusted (C)
	€million, annual		
Taxable income	829,281	829,281	900,744
	(8,150)	(8,150)	(10,527)
Self-employment income	105,313	105,313	176,775
	(4,008)	(4,008)	(7,798)
Self-employment income of self-employed households	83,714	83,714	155,177
	(3,734)	(3,734)	(7,464)
Taxable income of self-employed households	103,759	103,759	175,222
	(3,955)	(3,955)	(7,660)
IRPEF	165,823	165,823	191,615
	(2,772)	(2,772)	(3,838)
Self employed SSCs	22,477	22,477	31,141
	(629)	(629)	(869)
Pensions and other transfers	281,503	281,503	281,412
	(3,448)	(3,448)	(3,448)
Disposable income	665,927	737,390	702,839
	(5,332)	(7,832)	(6,416)

	Changes in % w.r.t. Reported	
Taxable income	0	8.62
Reported Self-employment income	0	67.86
IRPEF	0	15.55
Self employed SSCs	0	38.55
Pensions and other transfers	0	-0.03
Disposable income	10.73	5.54

Notes: IRPEF stands for national and regional personal income tax.
Standard errors in parentheses. Source: authors' elaboration based on EUROMOD (HBS-fiscal data).

Table 9: Average tax rates in different tax compliance scenarios, 2018

	True	Evasion-adjusted
1	0.206 (0.026)	0.284 (0.026)
2	0.173 (0.007)	0.226 (0.005)
3	0.173 (0.006)	0.244 (0.004)
4	0.187 (0.007)	0.271 (0.005)
5	0.203 (0.007)	0.298 (0.005)
6	0.213 (0.009)	0.314 (0.005)
7	0.220 (0.007)	0.339 (0.004)
8	0.222 (0.008)	0.374 (0.003)
9	0.223 (0.006)	0.387 (0.003)
10	0.226 (0.004)	0.454 (0.001)
All	0.213 (0.003)	0.368 (0.002)

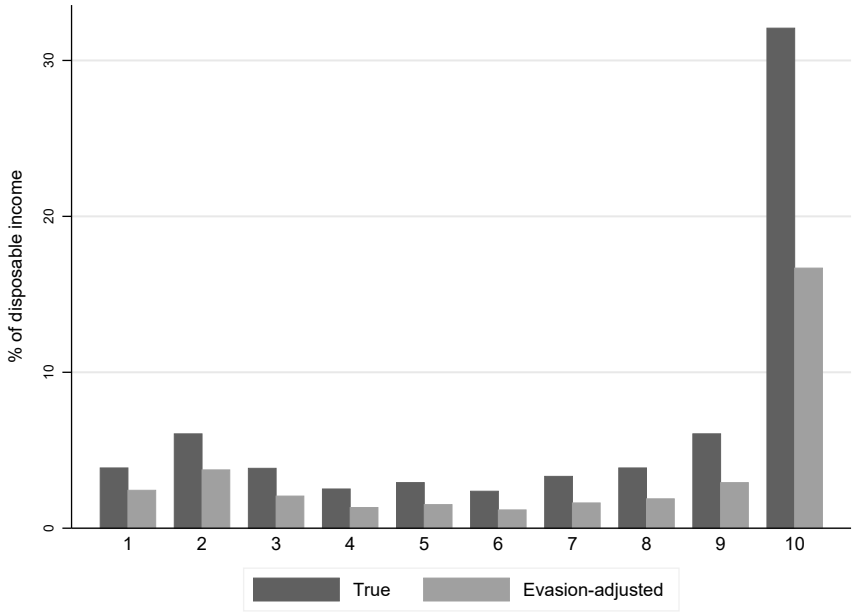
Notes: Averages of individual tax rates. Decile groups based on taxable income in the evasion adjusted scenario. Standard errors shown in parentheses. Source: authors' elaboration based on EUROMOD (HBS - fiscal data).

Table 10: Inequality and redistributive effects in different tax compliance scenarios, 2018

	Reported	True	Evasion-adjusted
Market income, Gini	0.556	0.576	0.576
	(0.002)	(0.003)	(0.003)
Taxable income, Gini	0.425	0.425	0.448
	(0.002)	(0.002)	(0.003)
Disposable income, Gini	0.348	0.380	0.361
	(0.002)	(0.003)	(0.003)
Reynolds-Smolenski	0.208	0.195	0.216
	(0.002)	(0.002)	(0.002)

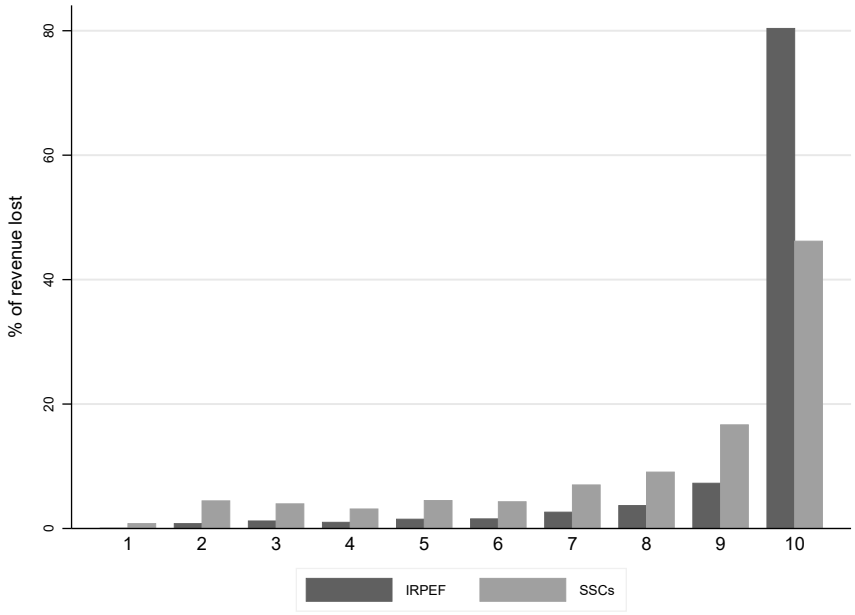
Notes: Standard errors in parentheses. Source: authors' elaboration based on EUROMOD (HBS-fiscal data).

Figure 2: Changes in disposable income in different tax compliance scenarios, 2018



Notes: Changes in disposable income in the two scenarios with respect to the observed scenario. Decile groups based on equivalized disposable income in the evasion-adjusted scenario. Source: authors' elaboration based on EUROMOD (HBS-fiscal data).

Figure 3: Distribution of revenue losses, 2018



Notes: Proportion of total loss revenue by decile group, based on equivalized disposable income in the evasion-adjusted scenario. Source: authors' elaboration based on EUROMOD (HBS-fiscal data).

7 Discussion and Conclusions

The increased availability of administrative data and their link with HBS has contributed to revive an established methodology for detecting traces of taxable income based on the estimation of food-expenditure reduced form equations. In fact, they now allow an estimation of tax evasion in countries where this methodology was unfeasible given the lack of an household income variable in the national HBS, such as Italy, and they reduce the attenuation bias due to a wide range of measurement errors, including recall errors of surveys, transitory components frequent for irregularly received self-employment income, unexpected gains and losses.

This dataset is novel for Italy and interesting on its own right because Italy has a large share of self-employed workers and a large share of under reporting when compared to other developed countries (Schneider et al. (2015)). Our findings, using a classification of self-employed households as those with at least 50% of their income from self-employment, suggest that self-employed workers underreport about 40% of their income to tax authorities, which is in line with the official estimates for Italy obtained by using National Accounts. It should,

however, be noted that the hidden income share reduces to around 30% if the usual 25% share was used instead for identifying self-employed units, which is similar to that found in the United States (Hurst et al., 2014) and only slightly larger than in the United Kingdom (Cabral et al., 2019) or in Sweden (Engström and Hagen, 2017). Similar conclusions with different a methodology was reached also by D'Attoma (2020) comparing Italy and the United States. These results suggest that the propensity to tax evasion is not so different in Italy among self-employed households, but is overall larger because the share of self-employed households is larger.

In contrast from top-down aggregate estimates of tax gap, which is the commonly used approach by tax administrations worldwide, our work is able to throw new lights on two disputed but unexplored, particularly in Italy, aspects of self-employment income tax evasion. We studied the heterogeneity of income underreporting among the self-employed and found that self-employed single households evade significantly more than in-couple ones. The exception is couples whose head is college educated, who can possibly exploit their skills or social networks to better coordinate with the spouse or reduce the risk of being caught.

Hence, we document the consequences of self-employed tax evasion on income inequality and redistribution. Once tax evasion is taken into account for reconstructing true self-employed income, we find that inequality levels, as measured by the Gini index of market and disposable income, are higher than in the absence of evasion information. This confirms Hurst et al. (2014) conclusion that neglecting self-employment underreporting would provide a large bias to standard statistics. Distributional analysis show that households at the top of the income distribution reap most of the gain by consistently reducing their tax burden in terms of average tax rates and total amount of taxes payed. We also find that the effects of self-employment income tax evasion on public budget are relevant, with a loss of about €26 billion of personal income tax and €8.6 billion in self-employed social security contributions, corresponding to about 16% and 39% of their respective total actual revenues.

Our work makes different contributions to the the existing literature. First, we ameliorate the expenditure-based method because we depart from the assumption of a constant underreporting share among all self-employed households, and we identify different underreporting shares within the category of self-employed households depending on some observable characteristics. Second, we largely reduce the extent of measurement error by using 7-year average income as permanent income and we further cleanse endogeneity of the income variable using the stock of financial wealth as an instrumental variable. Third, by using our finding of traces of hidden taxable income with a static tax-benefit microsimulation model, we study the extent to which the distribution of the tax burden and the effects of the tax-benefit system change when reconstructing the true income of a particular category of taxpayers, which contributes to the debate on the inequality of tax evasion (Alstadsæter et al., 2019; Nygård et al., 2018).

The results in this paper have several policy implications. Our evidence

supports the use of the expenditure-based methodology to complement top-down results based on national accounts to find traces of taxable income. The identification of differences in the underreporting behavior within the category of self-employed (i.e., heterogeneous fractions of underreported income) can be of direct interest for tax administrations in, at least, two directions. First, the evasion rates estimated at household level can be used to integrate the input data of tax-benefit microsimulation models so as to take into account tax evasion and true income distributions when analysing the revenue and distributional effects of tax-benefit reforms, such as minimum income programs and personal income tax reforms. Second, information on the heterogeneous features of tax evasion and the specific profiles of evaders can be useful when designing and conducting targeted tax audits, with a clear reduction of tax administration costs. Finally, we provide empirical ground to understand the complex inequality-evasion nexus (Bloomquist, 2003), and we support the adoption of policy programs aimed at tackling inequality through a reduction of tax evasion and an improvement of the distributional effects of the tax-benefit system.

Hopefully, we have convincingly proven that creativity in searching for traces of hidden taxable income can be rewarding.

8 Appendix

In this Appendix, we provide some further sensitivity checks to our main results.

In Table 11, we test the robustness of alternative ways to classify self-employed households by dropping type-W households who present some self-employment income, hence selecting only truly third-party-reporting households (Models 1 and 2) or by selecting households by the age group of the household head, using the 25-55 age range. The robustness of our main results are supported by these results.

In Table 12, we check the robustness of results presented to an alternative dependent variable, namely home utility expenditures, as in Cabral et al. (2019); Kukk et al. (2019); Albarea et al. (2019). We estimate a slightly lower percentage of undeclared income, with high statistical significance (Table 12), which overall confirms the robustness of our main results.

Finally, in Table 13, we present budgetary effects of different tax compliance scenarios similarly to Table 8 but assuming no heterogeneity of tax evasion behavior.

Table 11: Expenditure equations for (log) home utility expenditures and alternative measures of permanent income using different definitions of self-employed households, with sample selections.

	Model 1		Model 2		Model 3		Model 4	
	(A) OLS	(B) IV	(C) OLS	(D) IV	(E) OLS	(F) IV	(G) OLS	(H) IV
	Income = Pre-tax Total Family Income							
Self-employed	0.056***	0.069***	0.043***	0.056***	0.057***	0.069***	0.049***	0.058***
	(0.016)	(0.022)	(0.015)	(0.019)	(0.017)	(0.024)	(0.017)	(0.022)
Aver. inc. (7-year)	0.094***	0.143***	0.093***	0.153***	0.106***	0.149***	0.106***	0.148***
	(0.010)	(0.050)	(0.010)	(0.051)	(0.012)	(0.053)	(0.012)	(0.053)
$(1-\kappa)$	0.448***	0.383***	0.370***	0.306***	0.412***	0.370***	0.368***	0.327***
	(0.095)	(0.079)	(0.102)	(0.074)	(0.096)	(0.078)	(0.101)	(0.079)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.261	0.258	0.261	0.257	0.218	0.216	0.218	0.216
N. obs	16037	16037	16638	16638	12613	12613	12613	12613
N. obs Self-Employed	1767	1767	2368	2368	1437	1437	1939	1939
Share Self-Employed	0.866	0.866	0.742	0.742	0.866	0.866	0.739	0.739
F-stat		188.678		205.722		175.816		179.221
	Income = Post-tax Total Family Income							
Self-employed	0.057***	0.073***	0.044***	0.059***	0.057***	0.072***	0.049***	0.061***
	(0.016)	(0.023)	(0.015)	(0.020)	(0.017)	(0.025)	(0.017)	(0.023)
Aver. inc. (7-year)	0.098***	0.156***	0.098***	0.167***	0.112***	0.164***	0.111***	0.162***
	(0.012)	(0.055)	(0.011)	(0.056)	(0.013)	(0.058)	(0.013)	(0.058)
$(1-\kappa)$	0.439***	0.371***	0.363***	0.296***	0.402***	0.356***	0.357***	0.312***
	(0.092)	(0.074)	(0.099)	(0.068)	(0.093)	(0.073)	(0.098)	(0.074)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.26	0.256	0.26	0.255	0.216	0.214	0.216	0.214
N. obs	16037	16037	16638	16638	12613	12613	12613	12613
N. obs Self-Employed	1767	1767	2368	2368	1437	1437	1939	1939
Share Self-Employed	0.866	0.866	0.742	0.742	0.866	0.866	0.739	0.739
F-stat		184.09		200.355		182.797		186.995

Notes:
Model 1: A self-employed household has at least 50% of its income from self-employment. Households with positive amount of self-employment but amounting to less than 50% of their total income are dropped from the analysis.
Model 2: A self-employed a household with at least 25% of its income from self-employment. All families with some positive self-employment income but totaling to less than 25% are dropped from the analysis.
Model 3: A self-employed household has at least 50% of its income from self-employment. Only households with male head aged 25-55 are kept in the sample.
Model 4: A self-employed household has at least 25% of its income from self-employment. Only households with male head aged 25-55 are kept in the sample.
Controls include household head age and gender, in-couple dummy interacted with education (primary, secondary or tertiary) of the spouse, household size, a dummy for presence of kids, family consumption of sin goods, a full set of macro area of residence dummies, household head education and building property wealth (cadastral values).
Standar errors are adjusted for 109 clusters at the province of family residence.

Table 12: Home utility (log) expenditure equation with alternative measures of permanent income.

	Model 1		Model 2	
	(A) OLS	(B) IV	(C) OLS	(D) IV
	Income = Pre-tax Total Family Income			
Self-employed	0.053***	0.068***	0.055***	0.071***
	(0.020)	(0.024)	(0.017)	(0.022)
Aver. inc. (7-year)	0.099***	0.156***	0.094***	0.154***
	(0.010)	(0.045)	(0.009)	(0.048)
$(1-\kappa)$	0.415***	0.355***	0.441***	0.369***
	(0.113)	(0.086)	(0.098)	(0.075)
Controls	Yes	Yes	Yes	Yes
R-squared	0.214	0.211	0.263	0.259
N. obs	18198	18198	18198	18198
N. obs Self-Employed	1767	1767	1767	1767
Share Self-Employed	0.866	0.866	0.866	0.866
F-stat		207.798		207.798
	Income = Post-tax Total Family Income			
Self-employed	0.054***	0.072***	0.056***	0.074***
	(0.020)	(0.024)	(0.017)	(0.023)
Aver. inc. (7-year)	0.104***	0.171***	0.099***	0.168***
	(0.011)	(0.050)	(0.011)	(0.053)
$(1-\kappa)$	0.406***	0.344***	0.432***	0.357***
	(0.109)	(0.080)	(0.096)	(0.070)
Controls	Yes	Yes	Yes	Yes
R-squared	0.213	0.209	0.262	0.257
N. obs	18198	18198	18198	18198
N. obs Self-Employed	1767	1767	1767	1767
Share Self-Employed	0.866	0.866	0.866	0.866
F-stat		205.36		205.36

Notes:
Model 1: A self-employed household has at least 50% of its income from self-employment.
Model 2: A self-employed household has at least 25% of its income from self-employment.
Controls include household head age and gender, in-couple dummy interacted with education (primary, secondary or tertiary) of the spouse, household size, a dummy for presence of kids, family consumption of sin goods, a full set of macro area of residence dummies, household head education and building property wealth (cadastral values).
Standar errors are adjusted for 109 clusters at the province of family residence.

Table 13: Budgetary effects of different tax compliance scenarios assuming no heterogeneity in tax evasion as in Table 3. Values updated to year 2018

	Reported (A)	True (B)	Evasion-adjusted (C)
		€million, annual	
Taxable income	829,281	829,281	891,705
	(8,150)	(8,150)	(9,836)
Self-employment income	105,313	105,313	167,737
	(4,008)	(4,008)	(6,864)
Self-employment income of self-employed households	83,714	83,714	146,138
	(3,734)	(3,734)	(5,563)
Taxable income of self-employed households	103,759	103,759	166,183
	(3,955)	(3,955)	(2,775)
IRPEF	165,823	165,823	187,137
	(2,772)	(2,772)	(3,522)
Self employed SSCs	22,477	22,477	31,394
	(629)	(629)	(844)
Pensions and other transfers	281,503	281,503	281,380
	(3,448)	(3,448)	(3,448)
Disposable income	665,927	728,351	697,993
	(5,332)	(7,832)	(6,078)
		Changes in %	
Taxable income		0.00	7.53
Repported Self employment income		0.00	59.28
IRPEF		0.00	12.85
Self employed SSCs		0.00	39.67
Pensions and other transfers		0.00	-0.04
Disposable income		9.37	4.82

Notes: IRPEF stands for national and regional personal income tax.
Standard errors in parentheses. Source: authors' elaboration based on EUROMOD (HBS-fiscal data).

References

Albarea, A., Bernasconi, M., Marenzi, A., and Rizzi, D. (2019). Income Underreporting and Tax Evasion in Italy: Estimates and Distributional Effects. *Review of Income and Wealth*.

Alm, J. (1999). Tax Compliance and Administration. In Hildreth, W. B. and A.Richardson, J., editors, *Handbook on Taxation*, pages 741–768. New York, NY: Marcel Dekker, Inc.

Alstadsæter, A., Johannesen, N., and Zucman, G. (2019). Tax evasion and inequality. *American Economic Review*, 109(6):2073–2103.

Andreoni, J., Erard, B., and Feinstein, J. (1998). Tax Compliance. *Journal of Economic Literature*, 36(2):818–860.

Bernasconi, M. and Marenzi, A. (1997). *Ricerche quantitative per la politica economica 1997*, chapter Gli effetti redistributivi dell'evasione fiscale in Italia, pages 1 – 38. Bank of Italy, Rome.

Bloomquist, K. M. (2003). Tax evasion, income inequality and opportunity costs of compliance. In *Proceedings. Annual Conference on Taxation and Minutes of the Annual Meeting of the National Tax Association*, volume 96, pages 91–104. JSTOR.

Braiotta, A., Carfora, A., Pansini, R. V., and Pisani, S. (2020). Tax Gap and Redistributive Aspects across Italy. *Scienze Regionali, Italian Journal of Regional Science*, (2/2020):205–226.

Cabral, A. C. G., Gemmell, N., and Alinaghi, N. (2020). Are Survey-Based Self-Employment Income Under-Reporting Estimates Biased? New Evidence from Matched Register and Survey Data. *International Tax and Public Finance*.

Cabral, A. C. G., Kotsogiannis, C., and Myles, G. (2019). Self-Employment Income Gap in Great Britain: How Much and Who? *CESifo Economic Studies*, 65(1):84–107.

Ceriani, L., Figari, F., and Fiorio, C. (2020). EUROMOD Report. Italy 2014-2017. Technical report, Institute for Social and Economic Research, University of Essex.

Chung, C., Gibson, J., and Kim, B. (2010). CPI Mismeasurements and Their Impacts on Economic Management in Korea. *Asian Economic Papers*, 9(1):1–15.

CRA (2016). Tax Gap in Canada: A Conceptual Study. Technical report, Canadian Revenue Agency, Canada.

D'Attoma, J. (2020). More Bang For Your Buck: Tax Compliance in the United States and Italy. *Journal of Public Policy*, 40(1):1–24.

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

Feinstein, J. S. (1991). An econometric analysis of income tax evasion and its detection. *The RAND Journal of Economics*, pages 14–35.

Feldman, N. and Slemrod, J. (2007). Estimating tax noncompliance with evidence from unaudited tax returns. *Economic Journal*, 117(518):327–352.

Fellner-Röhling, G., Sausgruber, R., and Traxler, C. (2013). Testing enforcement strategies in the field: Threat, moral appeal and social information. *Journal of the European Economic Association*, 11(3):634–660.

Fiorio, C. V. and D'Amuri, F. (2005). Workers' Tax Evasion in Italy. *Giornale degli economisti e Annali di economia*, pages 247–270.

Gibson, J., Stillman, S., and Le, T. (2008). CPI Bias and Real Living Standards in Russia During the Transition. *Journal of Development Economics*, 87(1):140–160.

Gottschalk, P. and Moffitt, R. (1994). The Growth of Earnings Instability in the U.S. Labor Market. *Brookings Papers on Economic Activity*, 25(2):217–272.

HMRC (2019). Measuring tax gaps 2019 edition - tax gap estimates for 2017 to 2018. Technical report, HM Revenue and Customs, UK.

Hofmann, E., Voracek, M., Bock, C., and Kirchler, E. (2017). Tax Compliance Across Sociodemographic Categories: Meta-analyses of Survey Studies in 111 Countries. *Journal of Economic Psychology*, 62:63 – 71.

Hurst, E., Li, G., and Pugsley, B. (2014). Are Household Surveys Like Tax Forms? Evidence from Income Underreporting of the Self-Employed. *The Review of Economics and Statistics*, 96(1):19–33.

IRS (2019). Tax Gap Estimates for Tax Years 2011-2013. Technical report, Internal Revenue Service, US.

Johns, A. and Slemrod, J. (2010). The Distribution of Income Tax Noncompliance. *National Tax Journal*, 63(3):397–418.

Kim, B., Gibson, J., and Chung, C. (2017). Using Panel Data to Estimate Income Under-Reporting by the Self-Employed. *Manchester School*, 85(1):41–64.

Kleven, H. J., Knudsen, M. B., Kreiner, C. T., Pedersen, S., and Saez, E. (2011). Unwilling or unable to cheat? evidence from a tax audit experiment in denmark. *Econometrica*, 79(3):651–692.

Kukk, M., Paulus, A., and Staehr, K. (2019). Cheating in europe: Underreporting of self-employment income in comparative perspective. *International Tax and Public Finance*.

Levine, R. and Rubinstein, Y. (2017). Smart and illicit: Who becomes an entrepreneur and do they earn more? *The Quarterly Journal of Economics*, 132(2):963–1018.

Luttmer, E. F. P. and Singhal, M. (2014). Tax morale. *Journal of Economic Perspectives*, 28(4):149–68.

MEF (2019a). Analisi statistiche - Dichiarazioni 2019 - Anno d'imposta 2018. Technical report, Italian Ministry of the Economy and Finance. Accessed: 2020-05-20.

MEF (2019b). Relazione sull'economia non osservata e sull'evasione fiscale e contributiva – anno 2019. Technical report, Italian Ministry of the Economy and Finance. `http://www.dt.mef.gov.it/modules/documenti_it/analisi_progammazione/documenti_programmatici/def_2019/Allegato_NADEF_2019_Relazione_evasione_fiscale_e_contributiva.pdf`.

Nygård, O. E., Slemrod, J., and Thoresen, T. O. (2018). Distributional implications of joint tax evasion. *The Economic Journal*, 129(620):1894–1923.

Paulus, A. (2015). Income Underreporting Based on Income-expenditure Gaps: Survey vs Tax Records. ISER Working Paper Series 2015-15, Institute for Social and Economic Research.

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

Ricciuto, L., Tarasuk, V., and Yatchew, A. (2006). Socio-demographic influences on food purchasing among canadian households. *European journal of clinical nutrition*, 60:778–90.

Rondinelli, C. (2014). On the structure of italian households: Consumption patterns during the recent crises. *Politica economica*, 30(2-3):235–260.

Schneider, F., Raczkowski, K., and Mróz, B. (2015). Shadow economy and tax evasion in the eu. *Journal of Money Laundering Control*.

Slemrod, J. (2019). Tax compliance and enforcement. *Journal of Economic Literature*, 57(4):904–54.

Slemrod, J., Blumenthal, M., and Christian, C. (2001). Taxpayer Response to an Increased Probability of Audit: Evidence from a Controlled Experiment in Minnesota. *Journal of Public Economics*, 79(3):455–483.

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(1):25–53.

Slemrod, J. and Yitzhaki, S. (2002). Tax avoidance, evasion, and administration. In Auerbach, A. J. and Feldstein, M., editors, *Handbook of Public Economics*, volume 3, chapter 22, pages 1423–1470. Elsevier, 1 edition.

Solon, G. (1992). Intergenerational Income Mobility in the United States. *American Economic Review*, 82(3):393–408.

Stillman, S. and Thomas, D. (2008). Nutritional Status During an Economic Crisis: Evidence from Russia. *Economic Journal*, 118(531):1385–1417.

Sutherland, H. and Figari, F. (2013). EUROMOD: The European Union Tax-benefit Microsimulation Model. *International Journal of Microsimulation*, 1(6):4–26.

Tedds, L. M. (2008). Estimating the income reporting function for the self-employed. *Empirical Economics*, 38(3):669–687.

Turrell, G., Hewitt, B., Patterson, C., and Oldenburg, B. (2003). Measuring socio – economic position in dietary research: Is choice of socio – economic indicator important? *Public Health Nutrition*, 6(2):191 – 200.

Worsley, A., Blasche, R., Ball, K., and Crawford, D. (2004). The relationship between education and food consumption in the 1995 australian national nutrition survey. *Public Health Nutrition*, 7:649–63.