The IRS Research Bulletin

Proceedings of the 2020 IRS / TPC Research Conference

Research, Applied Analytics & Statistics

Publication 1500 (Rev. 5-2021) Catalog Number 11546J Department of the Treasury **Internal Revenue Service** www.irs.gov

Papers given at the

10th Annual Joint Research Conference on Tax Administration

Cosponsored by the IRS and the Urban-Brookings Tax Policy Center

Held Virtually
June 18, 2020

Compiled and edited by Alan Plumley*
Research, Applied Analytics, and Statistics, Internal Revenue Service

*Prepared under the direction of Barry W. Johnson, Acting IRS Chief Research and Analytics Officer

Foreword

This edition of the *IRS Research Bulletin* (Publication 1500) features selected papers from the IRS-Tax Policy Center (TPC) Research Conference held virtually on June 18, 2020. Conference presenters and attendees included researchers from many areas of the IRS, officials from other Government agencies, and academic and private sector experts on tax policy, tax administration, and tax compliance. Many people participated in this, our first fully virtual conference. Videos of the presentations are archived on the Tax Policy Center website to enable additional participation. Attendees participated in the discussions by submitting questions via e-mail as the sessions proceeded.

The conference began with welcoming remarks by Eric Toder, Co-Director of the Tax Policy Center, and by Barry Johnson, the Acting IRS Chief Research and Analytics Officer. The remainder of the conference included sessions on behavioral responses to audits, new insights on taxpayer behavior, advances in taxpayer service, and doing more with less. The keynote speaker was former IRS Commissioner Charles Rossotti, who explained his proposals for improving tax compliance.

We trust that this volume will enable IRS executives, managers, employees, stakeholders, and tax administrators elsewhere to stay abreast of the latest trends and research findings affecting tax administration. We anticipate that the research featured here will stimulate improved tax administration, additional helpful research, and even greater cooperation among tax administration researchers worldwide.

Acknowledgments

This IRS-TPC Research Conference was the result of preparation over a number of months by many people. The conference program was assembled by a committee representing research organizations throughout the IRS. Members of the program committee included: Alan Plumley, Brett Collins, Tom Hertz, Lindsay Schrock, and John Weaver (RAAS); Fran Cappelletti (Taxpayer Advocate); Melissa Hayes (Large Business & International Division); Quinton Anderson and Alexis Kinikin (Small Business / Self-Employed Division); Alcora Walden (Online Services); and Rob McClelland (Tax Policy Center). In addition, Ann Cleven and Hailey Roemer from the Tax Policy Center oversaw numerous details to ensure that the conference ran smoothly.

This volume was prepared by Camille Swick and Lisa Smith (layout and graphics) and Beth Kilss and Georgette Walsh (editors), all of the IRS Statistics of Income Division of RAAS. The authors of the papers are responsible for their content, and views expressed in these papers do not necessarily represent the views of the Department of the Treasury or the Internal Revenue Service.

We appreciate the contributions of everyone who helped make this conference a success.

Barry Johnson
Acting IRS Chief Research and Analytics Officer

10th Annual IRS-TPC Joint Research Conference on Tax Administration

Contents

4. Doing More With Less

5. Appendix

2

New Insights on Taxpayer Behavior

Bazzoli ◆ Di Caro ◆ Figari ◆ Fiorio ◆ Manzo

An Analysis of Self-Employed Income Tax Evasion in Italy With a Consumption-Based Methodology

Martina Bazzoli, Paolo Di Caro, and Marco Manzo (Italian Ministry of Economy and Finance), Francesco Figari (University of Insubria), and Carlo Fiorio (University of Milan)

1. Introduction

The study of personal income tax evasion and individual underreporting is important, among other factors, for knowing the true income distribution in a given country and for providing more accurate evaluations of the redistributive effects of tax policies (Matsaganis e*t al.* (2010)). This is particularly relevant in countries like Italy, where tax evasion is high in comparison to other developed countries and it shows persistence across time (Schneider e*t al.* (2015)): in 2018, the Italian personal income tax (PIT) gap was equal to about 31.5 million euro, one-third of PIT revenues for the same year (Ministry of Economy and Finance (2020)). Measuring personal income tax evasion, however, is not a trouble-free task given the invisible nature of evasion activities and the need of having detailed information on individuals (Slemrod and Weber (2012)).

In some countries, such as the United Kingdom, the United States, and Denmark, the availability of administrative micro data based on random tax audits provides good information that can be used for estimating personal income tax evasion with a bottom-up approach. Alternative bottom-up techniques have been used in those countries where survey and tax data can be merged, either statistically and/or exactly through personal identification codes: discrepancies methods (Paulus (2015)), and expenditure-based analyses (Hurst e*t al.* (2014); Cabral e*t al.* (2019)). In Italy, due to the lack of random tax audits and the unavailability of tax microdata until now, income tax evasion has been mainly estimated by using the top-down approach that combines aggregate information on national accounts and tax data. In this country, bottom-up applications have been applied for research purposes (Bernasconi and Marenzi (1997); Fiorio and D'Amuri (2005)), with renewed interest in recent years (Albarea e*t al.* (2019); Lalla e*t al.* (2019)). In the next section, which contains the literature review, we discuss the added value of using bottom-up approaches for analyzing personal income tax evasion in countries such as Italy where the top-down methodology is the only one available.

In this work, for the first time for Italy, we study self-employed personal income tax evasion by applying the bottom-up approach that relies on the consumption-based methodology (Pissarides and Weber (1989)). Specifically, we build a novel dataset based on the exact matching of tax administrative microdata from individual tax declarations over the period 2010–2016 with information from the Italian Household Budget Survey (HBS) for the year 2013 that does not contain income variables. The exact matching of income and consumption data, which has been conducted by the IT Department of the Ministry of Economy and Finance (MEF) to preserve anonymity, allows us to rule out the issues that are present when adopting statistical matching techniques (Atkinson and Brandolini (2001)). Moreover, the availability of panel data regarding income covering 7 years gives us the possibility of overcoming problems related to the usage of current income in the estimation of the consumption-income curves (Engström and Hagen (2017)).

The second contribution of our study is to provide evidence on the heterogeneity of the estimates of self-employed income tax evasion in Italy. Specifically, we start by investigating the different evasion rates of the self-employed across the Italian macroareas (North, Centre, South), which is justified by the relevant territorial economic and social differences that are present in Italy, which can have consequences on the tax evasion behaviour (D'Attoma (2019)). One of the policy implications of such results is that we support possible region-specific tax compliance actions. In addition, we depart from the aggregate definition of self-employed, and we make a distinction between small entrepreneurs and liberal professionals (e.g., lawyers, doctors, accountants,

etc.). This separation can be made thanks to our administrative data that allows us to identify the particular category of self-employed under analysis. From an economic point of view, recent evidence suggests that entrepreneurs can show different characteristics (i.e., risk profile, education, etc.) than the rest of self-employed workers (Levine and Rubinstein (2017)). From a policy perspective, the knowledge of differences in tax evasion rates within the category of self-employed is important to better tailoring policies aimed at reducing tax evasion.

Our results, which are robust to alternative consumption and income variables, and remain valid after comparing Ordinary Least Squares (OLS) and Instrumental Variable (IV) estimates, suggest that the underreporting gap of self-employed households ranges from 27 percent to 35 percent. These findings are not significantly dissimilar to the results obtained by applying the same methodology to other institutional contexts such as the United States (Hurst *et al.* (2014)), and the United Kingdom (Cabral *et al.* (2019)). Interestingly, this result supports, in contrast to the popular wisdom, the recent experimental evidence suggesting that the extent of tax evasion in Italy is not so different from that registered in other countries (D'Attoma *et al.* (2017)). In addition, we find that self-employed households located in the North of the country evade more income, relative to dependent-worker households living in the same macroarea, than in the rest of the country. Also, we document that liberal professionals underreport a share of income that is about twice that underreported by small entrepreneurs.

The rest of the work is organized as follows: The next section overviews the related literature. Then, we present the data and the methodology. The fourth section contains the results. The final section concludes with some policy implications.

2. Literature Review

There are two approaches commonly used for quantifying personal income tax evasion: top-down and bottom-up. The top-down approach is used by tax administrations where good microdata are not available and/or not accessible, and relies upon aggregate comparisons between national account data, which generally include evasion, and information collected by tax authorities, based on reported income only. There are some advantages in using the top-down approach. First, it provides time-series estimates of tax evasion. Second, it allows for the separation of gross and net tax gap, the latter taking into account the effects of tax compliance policies. Third, this approach does not request the availability of and the access to microdata. Yet, the top-down method presents the following shortcomings: It is not possible to disaggregate tax gap for different categories of taxpayers; and, it does not permit the study of the distributional effects of tax evasion in detail. For a more detailed discussion and an application to Italy, see Braiotta *et al.* (2020).

The bottom-up approach uses different sources of microdata and includes three different methods. The first method uses information derived from individual tax audits for approximating true income and calculating tax evasion. This method is typically applied in countries where random audits are available (United Kingdom, United States, and Denmark), and it requests the adoption of statistical corrections (e.g., uplift factor) for extending the results obtained for the used sample to the whole population (Clotfelter (1983); Feinstein (1991); Kleven *et al.* (2011)). This bottom-up method is able to provide time-series data on tax evasion; the main shortcoming is the cost of setting up random enquiry programs where they are not available.

The second method is based on the comparison of income data deriving from individual surveys and aggregate administrative data, on the general idea that surveys provide larger aggregate taxable income than administrative data, and assuming that taxpayers declare a closer-to-true income in an anonymous interview than in tax forms (Fiorio and D'Amuri (2005); Paulus (2015); Albarea *et al.* (2019)). This method, called the discrepancy approach, relies on the assumption that survey data are without errors and/or survey errors can be managed by the researcher in order to use income declared in surveys as true income (Koijen *et al.* (2014)). Moreover, given that surveys are usually available as repeated cross-sections, this method does not allow one to provide time-series estimates of tax evasion.

The third method is based on the comparison of income and consumption data for particular categories of taxpayers. Specifically, the so-called consumption-based method (Pissarides and Weber (1989)) relies upon

the estimation of expenditure curves for different groups of taxpayers with different underreporting possibilities, such as self-employed versus dependent workers, to approximate income tax evasion by the former relative to that of the latter. This method requires using as a consumption variable a set of items that—after controlling for observable characteristics—are assumed to be independent of selected groups, such as food. This methodology was first applied in the UK (Pissarides and Weber (1989)), and later applied in several other countries (Kukk *et al.* (2020)), including the United States (Hurst *et al.* (2014)), Canada (Tedds (2010)), and Sweden (Engström and Hagen (2017)). The consumption-based method requires the availability of detailed microdata, and the solution of some empirical issues such as: i) the choice of a good measure of permanent income; ii) the selection of consumption variables that does not conditionally depend on taxpayer occupations; iii) the matching between survey and administrative data, with statistical matching producing additional noise in the estimates. Moreover, this method does not allow for the production of time-series data of tax evasion given that it is usually based on cross-section survey collection. In the next sections, we discuss the application of the consumption-based method to the Italian case, and how we dealt with the practical issues in our case.

Despite the presence of some data and methodological problems, bottom-up estimates of tax evasion have recently regained importance among researchers and policymakers given the progressive accessibility to administrative microdata (Card *et al.* (2010)). In particular, bottom-up methods allow for integrating top-down estimates in several ways, particularly in those countries like Italy where bottom-up estimates are not generally used for policymaking. First, bottom-up results are able to integrate top-down findings, by providing robustness checks to the calculations obtained by using aggregate data. Second, the adoption of bottom-up methods allows for the identification of heterogeneous profiles of tax evasion based on individual and/or household characteristics. This can be particularly helpful for profiling tax evaders and supporting the design of more tailored tax audit policies. Third, microestimates of tax evasion used in combination with tax-benefit microsimulation models are important for throwing new light on the distributional implications of underreporting activities. For a discussion on the value-added of bottom-up results applied to Italy, see MEF (2020).

3. Data and Methodology

3.1 Data description

We use a novel consumption-income dataset for a representative sample of Italian households by linking the 2013 Italian Household Budget Survey (HBS), which is provided by the Italian National Institute of Statistics (ISTAT) on a yearly basis, with data on individual tax returns, available at the MEF, for the years 2010-2016. The HBS provides detailed information on consumption expenditures, with data on about 300 consumption items, and household characteristics (number of children, education of parents, age profiles, etc.) for about 20,700 households corresponding to about 50,000 individuals (Rondinelli (2014)). Unfortunately, and differently from other countries, the Italian HBS does not contain information on household income. 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 2 weeks.[1]

In this study, we use administrative information deriving from individual tax returns for measuring household income. Moreover, we employ individual and household characteristics present in tax returns for having a large set of observables. Administrative data allow for the measurement of the stock of property wealth at cadastral values that we use as an additional control variable. The panel structure of fiscal data allows us to construct a measure of declared individual income from year t-3 to year t+3, where t=2013, which is the year of the HBS, providing a good proxy of permanent income over a 7-year period. This implies that our results with the adoption of the permanent income proxy rule out the issues related to asymmetric income fluctuations among taxpayer categories that are present when using a measure of current income only (Engström and Hagen (2017)).

[1] The use of food consumption as dependent variable is motivated by the fact that food expenditures are usually uncorrelated with the self-employment status of a household, holding constant all other observable characteristics (Pissarides and Weber (1989)). Other contributions used different consumption items, available in the surveys, such as home utilities and health expenditures (Albarea *et al.* (2019)). In our data, we have also information on these additional consumption expenditures. Results with different dependent variables, available upon request, confirm the main findings of our work.

Although the merge between income (administrative) and consumption (survey) data is performed at the individual level, given that in Italy tax declarations are made individually, we perform our analysis at the household level. In this study, in line with the international literature following the initial contribution of Pissarides and Weber (1989), we define self-employed households as those households whose total income from self-employment is at most equal to 25 percent of total household income. In a companion work (Bazzoli *et al.* (2020)), we defined a household as self-employed if 50 percent of its income comes from self-employment. This choice is not without implications in terms of the aggregate consequences of self-employment underreporting that are sensitive to the particular definition of self-employed (Hurst *et al.* (2014)). It is worth noticing that our classification of self-employed households allows for the detection of about 12 percent of the total sample as self-employed, a share that is close to the total share of self-employed workers in the tax records. Our results are robust to the alternative classification of self-employed households including the self-declared status in a survey.

Table 1 shows some descriptive statistics for the whole sample, the share of self-employed households defined as those earning at least 25 percent of income from self-employment and the remaining ones, defined as dependent workers, which also include pensioners. Food expenditures (in logs) are higher for self-employed than for the rest of the population, while differences in declared household income are less marked notwithstanding the definition of income that is adopted (e.g., pre- and post-tax income, current vs 7-year average). These preliminaries, which are in line with the evidence for the U.S. (Hurst *et al.* (2014)), suggest that the income-consumption relations among categories of taxpayers shall be further investigated, as we will do in the next pages. Observe that, moreover, self-employed households are younger, mostly concentrated in the North of Italy, and headed by males, in comparison to dependent workers.

3.2 Methodology

To investigate the underreporting (tax evasion) rate of self-employed households, in comparison to the income reported by dependent worker households, we use the following consumption-income relationship:

Our dependent variable is the (log of) household food consumption lnC_i, where i denotes a given household; our main income variable is the (log of) household income declared in tax returns over the years 2010–2016. We call this measure a proxy of permanent income (Engström and Hagen (2017)). We also use an alternative income variable the (log of) household income declared in tax returns in 2013, the same year of the HBS survey used in this study, in order to provide a measure of current income.

$$lnC_i = \beta lnY_i + \mathbf{X}'\alpha + \gamma SE_i + \varepsilon_i \tag{1}$$

The set of baseline controls $\mathbf{X}'$ 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 the macroarea of residence dummies. Additional controls include also household head education and building property wealth (cadastral values). The controls, which are common in this literature (Cabral e*t al.* (2019)), are introduced to estimate the Engle curve conditional to the same individual and household characteristics for different categories of taxpayers, namely self-employed versus dependent workers. Specifically, the controls are used to rule out the influence of possible observable differences between categories of taxpayers in the investigated relationship.

The covariate SE_i is a dummy variable that takes value 1 for a given self-employed household, which we define as those households whose total income from self-employment is at least equal to 50 percent of total household income. The term ε_i is the error term of relation (1). The share of underreported income of self-employed households can be calculated as follows:

$$1 - \hat{k}_{SE} = 1 - exp\left[-\frac{\hat{\gamma}_{SE}}{\hat{\beta}}\right]. \tag{2}$$

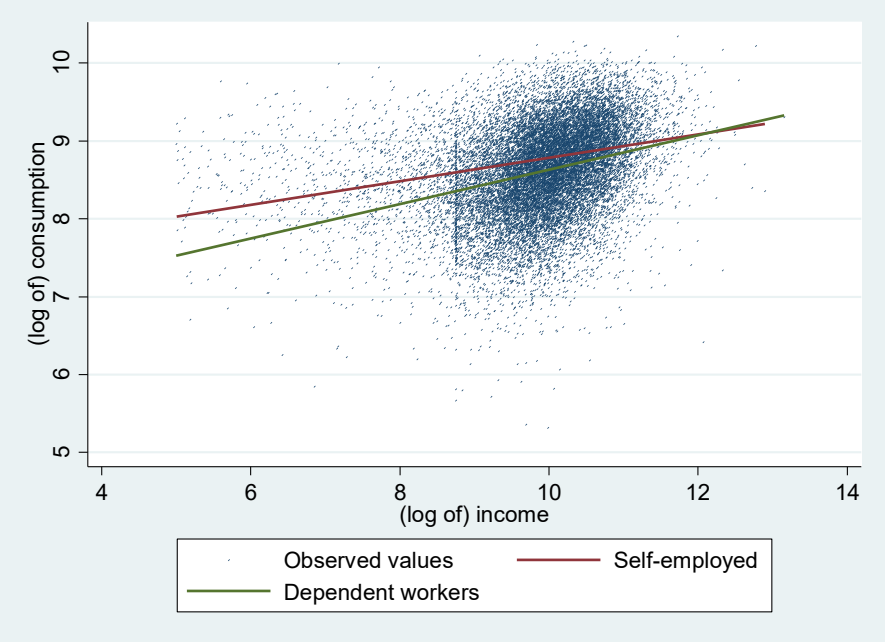
TABLE 1. Descriptive Statistics

	Whole sample	Self-employed	Dependent workers
log food expenditures	8.622	8.768	8.606
	(0.004)	(0.014)	(0.005)
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)
% 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)
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)
Sample size	18,198	1,767	16,431

NOTES: Our calculation is based on the selected sample; standard errors in parentheses. Self-employed households are identified as those with self-employment income equal to or larger than to 50 percent of total household income.

Relation (2) describes the proportion of unreported income of self-employed households ($1 - \hat{k}_{SE}$). It derives from the underlying assumption that self-employed households misreport their income, which is not third-party reported as in the case of dependent worker households, by a factor k, namely $Y_i^T = K_i Y_i^R$, with $K_i \leq 1$ where Y_i^T and Y_i^R denote true and reported income, respectively. For dependent workers, by assumption, $Y_i^T = Y_i^R$ and $k_i = 1$. Note that, in this approach, the factor k_i is assumed to be different among categories of households (i.e., self-employed vs dependent workers), but constant within the same category. In a different contribution, we relax this assumption by allowing for the possibility of having heterogeneous values for the factor k_i (Bazzoli *et al.* (2020)).

FIGURE 1. Income-Consumption Relation, Preliminary Evidence



NOTE: The graph reports the estimated values of the relation in (1) by applying the OLS estimator, when self-employed households are defined as having at most 25 percent of their total income from self-employment. The red line shows the predictions for self-employed (when the dummy SEi =1), while the green line shows the predictions for dependent workers (when the dummy SEi =0).

The graph in Figure 1 provides an illustration of the methodology that we use in this paper. It reports the values of the income-consumption relationship (dots), as estimated from the relation in (1). The red and green lines show the predicted values for self-employed and dependent workers, respectively. Two aspects are worth commenting upon. The predicted values for self-employed households are above those observed for dependent worker households, by suggesting that, for the same level of declared income, self-employed households have higher food expenditures than dependent workers. This difference, which is conditional to the same individual and household characteristics, can imply that self-employed households underreport the extent of their declared income, by denoting the presence of tax evasion. We are interested in quantifying the share of such underreporting that can be approximated by relation (2). Lastly, it is important to remember that we assume that dependent workers do not underreport their income, which can be restrictive particularly for private dependent workers (Paulus (2015)). If dependent workers can also misreport their income, our estimates of the tax evasion by self-employed households can be interpreted as a lower bound of the true level of tax evasion for such a category.

4. Results

4.1 Self-employed income tax evasion in Italy

In Tables 2 and 3, we report the estimates of the relation (1), and the estimated values of relation (2) reported in the tables as evasion rates, with the adoption of pre- and post-tax income, respectively. Using after-tax income, although subject to its own measurement issues, allows us to check to what extent fewer taxes paid by self-employed are allocated to consumption (Hurst *et al.* (2014)). We use both current and permanent income definitions in order to see how results change when smoothing income fluctuations with the adoption of the proxy of permanent income. For expositional convenience, we show the estimated coefficients of the self-employed dummy and income variables only. Estimates are obtained by clustering the errors at a provincial level for the 109 Italian provinces that describe the residence of the family.

In specifications (A-B), we use no controls, namely log consumption is regressed on a constant, the self-employment dummy, and the log of income. The specifications (C-D) include the set of controls, that is, 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, and family consumption of sin goods. The specifications (A-D) are obtained by applying OLS techniques. In the last two specifications (E-F), we apply the Instrumental Variable (IV) strategy, according to the existing literature since Pissarides and Weber (1989), where we use as an instrument the building property wealth measured using cadastral values. The IV strategy is useful for dealing with the endogeneity of current income in relation (1) and, moreover, for limiting measurement errors in the 7-year average income measure of permanent income (Engström and Hagen (2017)). The model diagnostics confirm the robustness of our findings.

Our results suggest that self-employed households consume on average more than 5 percent of what dependent worker households consume. The elasticity of consumption estimates suggest that changes in current income affect less than changes in the 7-year average income, consistently, with an interpretation of the latter as a better measure of permanent income. As for tax evasion, and when considering average income, we find that the underreporting gap of self-employed households ranges from 26 percent (specification F) to 35 percent (specification (D) when using the definition of after-tax family income (Table 3). The results are similar when using the definition of pre-tax family income, as in Table 2. Interestingly, such results are not significantly different from the findings obtained by applying the same methodology to other countries such as the United States (Hurst *et al.* (2014)), and the United Kingdom (Cabral *et al.* (2019)). In a different work (Bazzoli *et al.* (2020)), we showed that the average self-employment income tax evasion rate that we find here derives from heterogeneous underreporting shares that depend on specific individual and family characteristics (e.g., singles vs couples, age and educational levels, etc.).

TABLE 2. Self-Employment Income Tax Evasion, Pre-Tax Total Family Income

	(A)	(B)	(C)	(D)	(E)	(F)
	OLS	OLS	OLS	OLS	IV	IV
Self-employed	0.187***	0.177***	0.053***	0.055***	0.091***	0.083***
	(0.017)	(0.017)	(0.016)	(0.017)	(0.017)	(0.016)
Current income	0.197***		0.076***		0.201***	
	(0.009)		(0.008)		(0.022)	
Average income (7-yr)		0.233***		0.094***		0.201***
		(0.009)		(0.009)		(0.022)
Evasion rate	0.612***	0.534***	0.501***	0.441***	0.363***	0.340***
	(0.038)	(0.037)	(0.107)	(0.098)	(0.057)	(0.059)
Controls	No	No	Yes	Yes	Yes	Yes
R-squared	0.098	0.116	0.261	0.263	0.235	0.248
N. observations	18,198	18,198	18,198	18,198	18,198	18,198
N. obs self-employed	1,767	1,767	1,767	1,767	1,767	1,767
Share self-employed	0.775	0.775	0.775	0.775	0.775	0.775
F-stat					982.16	951.13

***Significant to the 1% level.
NOTE: 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). Standard errors are adjusted for 109 clusters at the province of family residence.

TABLE 3. Self-Employment Income Tax Evasion, Post-Tax Total Family Income

	(A)	(B)	(C)	(D)	(E)	(F)
	OLS	OLS	OLS	OLS	IV	IV
Self-employed	0.189***	0.183***	0.052***	0.056***	0.095***	0.089***
	(0.017)	(0.017)	(0.016)	(0.017)	(0.017)	(0.017)
Current income	0.216***		0.077***		0.222***	
	(0.011)		(0.009)		(0.024)	
Average income (7-yr)		0.259***		0.099***		0.223***
		(0.010)		(0.011)		(0.024)
Evasion rate	0.584***	0.506***	0.492***	0.432***	0.348***	0.329***
	(0.038)	(0.035)	(0.107)	(0.096)	(0.053)	(0.055)
Controls	No	No	Yes	Yes	Yes	Yes
R-squared	0.098	0.119	0.26	0.262	0.234	0.247
N. observations	18,198	18,198	18,198	18,198	18,198	18,198
N. obs self-employed	1,767	1,767	1,767	1,767	1,767	1,767
Share self-employed	0.775	0.775	0.775	0.775	0.775	0.775
F-stat					950.121	921.893

***Significant to the 1% level.
NOTE: 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). Standard errors are adjusted for 109 clusters at the province of family residence.

4.2 Regional distribution of self-employed tax evasion

In Italy, one of the most relevant dimensions of inquiry for analysing economic issues is represented by geography, given the long-lasting economic and social differences between the North and the South of the country. Such territorial differences produce several effects, including implications on inequality (Fiorio (2011); Di Caro (2017)), the distribution of evasion (Carfora e*t al.* (2018)), and the concentration of informal occupations (Di Caro and Sacchi (2020)). Understanding the region-specific patterns of self-employed tax evasion, a novelty of our contribution, is relevant because it provides further information on the concentration of evasion activities across the space (Wiseman (2013)), and, most importantly, it throws light into the regional distribution of tax revenues within the same country (González-Fernández and González-Velasco (2014)).

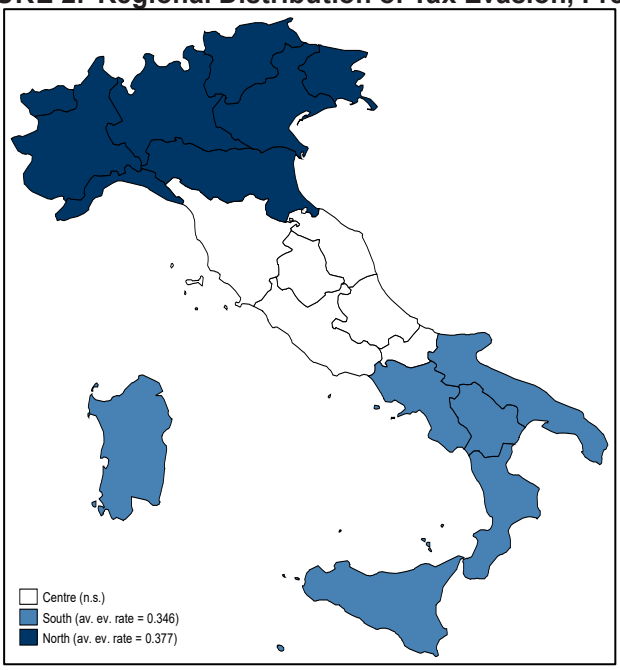
Our administrative data matched with the HBS consumption data allow for the analysis of the regional aspects of self-employed tax evasion, by providing a good sample size from a regional perspective. To keep a significant number of observations, however, we have preferred to produce estimates based on the three Italian macro-areas (North, Centre, South), which are obtained by aggregating the twenty Italian regions. In particular, we have estimated the relation (1) for each macro-area sub-sample separately. The results that we have obtained can be interpreted as the tax evasion rate of self-employed households compared to dependent workers households living in the same macro-area. In Figures 2 and 3, we report the shares of underreported income, as defined in the relation (2), for each macro-area when the income variable is pre- and post-tax household income, respectively. We have used the results obtained from the estimates of specification (F), with the IV strategy and all the control set, as discussed in the previous section. High self-employment evasion rates are marked in dark blue.

Some comments are worth discussing. We find that self-employed households underreport income relatively to dependent workers households located in the same area more in the regions located in the North (37 percent of their income) than in the rest of country. Indeed, in the South we detect a share of income underreported by self-employed equals to about 34 percent, while for the sub-sample of taxpayers located in the

[2] The lack of statistical significance for the analysis restricted to the sample of households located in the Centre can be due, among other factors, to the relatively lower number of observations in this sample (less than 18 percent of total observations). Moreover, in this macro-area, the income-consumption differences between self-employed and dependent workers households, conditional to other covariates, are very limited, possibly because there is the Lazio region where the Italian capital Rome is located and the category of dependent workers is the majority.

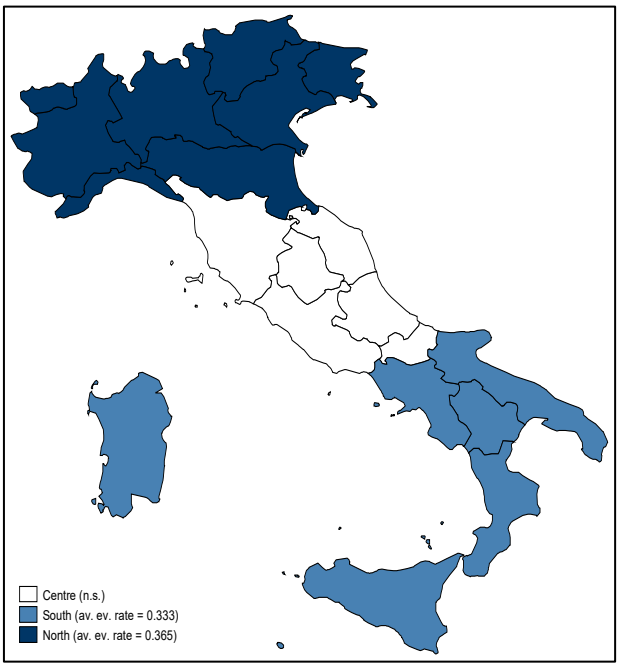
Centre we do not find significant results.[2] There are different reasons that can explain the higher evasion rates of self-employed registered in the Northern regions, which we left for future research. Note that, for instance, in this study we are not able to cover informal self-employed occupations that do not fill tax returns, which are possibly higher in the South than in the rest of the country (Di Caro and Sacchi (2020)).

FIGURE 2. Regional Distribution of Tax Evasion, Pre-Tax Income



NOTE: the graph shows the regional distribution of estimates in (1) by applying the specification (F), as in Table 2, for the sub-samples covering the three Italian macro-areas (North, Centre, South) separately; self-employed households are defined as having at least 50 percent of their total income from self-employment

FIGURE 3. Regional Distribution of Tax Evasion, Post-Tax Income



NOTE: the graph shows the regional distribution of estimates in (1) by applying the specification (F) as in Table 3 for the sub-samples covering the three Italian macro-areas (North, Centre, South), separately; self-employed households are defined as having at least 50 percent of their total income from self-employment.

4.3 Estimates for small entrepreneurs

There has been recent empirical evidence, particularly for the United States, on the fact that defining different categories of self-employed workers in a single way can produce misleading results (Levine and Rubinstein (2017)). Specifically, small and medium entrepreneurs, which are often classified as self-employed for the lack of detailed data, show significant differences in risk attitudes, organizational abilities, financial constraints and other economic and social traits, in comparison to the rest of self-employed workers (Levine and Rubinstein (2018)). Due to the lack of adequate data, to our knowledge, the consumption-income method has been applied in different countries by treating self-employed as a single category of workers. This has important policy implications since different types of self-employed can show different attitudes towards tax evasion and, most importantly, they need different tax compliance strategies. For instance, the introduction of compulsory electronic invoicing can be a good strategy for increasing tax compliance of small entrepreneurs in business-to-business (B2B) transactions, but not a sufficient tool for liberal professionals that are mostly involved in business-to-consumer (B2C) transactions. Therefore, throwing light into the different evasion profiles within the category of self-employed is necessary for guiding policymakers and, in particular, to clarify the distinction between the contrast to tax evasion in B2B transactions, which is due to omission to declare, and that in B2C transactions, which is more related to omission to invoice.

The tax return data that we use in this study gives us the possibility of making a distinction within the category of self-employed, by identifying small entrepreneurs (e.g., shop vendors, individual service firms). In this section, we have estimated the relation (1) for this category of self-employed households. The results that we have obtained can be interpreted as the tax evasion rate of small entrepreneurs households compared to dependent workers households. In Tables 4 and 5, we show the findings obtained for small entrepreneurs. Interestingly, our results, which are robust to alternative specifications and definition of the income variable, suggest that the share of income underreported by small entrepreneurs' households, relatively to dependent workers households, is lower than that registered for the entire category of self-employed households, namely 27 percent vs 34 percent. This difference, which needs further investigation on the reasons behind it, suggests the adoption of different compliance strategies with different costs for the tax administration, when trying to improve the compliance of self-employed.

TABLE 4. Income Tax Evasion, Small Entrepreneurs, Pre-Tax Total Family Income

	(A)	(B)	(C)	(D)	(E)	(F)
	OLS	OLS	OLS	OLS	IV	IV
Self-employed	0.177***	0.174***	0.034**	0.038**	0.068***	0.065***
	(0.017)	(0.017)	(0.016)	(0.016)	(0.017)	(0.017)
Current income	0.197***		(0.016)		0.203***	
	(0.010)		(0.016)		(0.021)	
Average income (7-yr)		0.233***		0.093***		0.209***
		(0.009)		(0.009)		(0.024)
Evasion rate	0.594***	0.526***	0.369***	0.336***	0.285***	0.276***
	(0.038)	(0.037)	(0.132)	(0.114)	(0.057)	(0.059)
Controls	No	No	Yes	Yes	Yes	Yes
R-squared	0.094	0.113	0.261	0.262	0.236	0.248
N. observations	18,198	18,198	18,198	18,198	18,198	18,198
N. obs self-employed	1,305	1,305	1,305	1,305	1,305	1,305
Share self-employed	0.769	0.769	0.769	0.769	0.769	0.769
F-stat					961.237	944.079

***Significant to the 1% level.

NOTE: 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). Standard errors are adjusted for 109 clusters at the province of family residence.

TABLE 5. Income Tax Evasion, Small Entrepreneurs, Post-Tax Total Family Income

	(A)	(B)	(C)	(D)	(E)	(F)
	OLS	OLS	OLS	OLS	IV	IV
Self-employed	0.176***	0.176***	0.033**	0.039**	0.070***	0.070***
	(0.017)	(0.017)	(0.016)	(0.016)	(0.017)	(0.017)
Current income	0.215***		0.076***		0.224***	
	(0.011)		(0.009)		(0.024)	
Average income (7-yr)		0.259***		0.097***		0.224***
		(0.010)		(0.011)		(0.024)
Evasion rate	0.560***	0.494***	0.356***	0.329***	0.267***	0.269***
	(0.038)	(0.036)	(0.133)	(0.111)	(0.053)	(0.054)
Controls	No	No	Yes	Yes	Yes	Yes
R-squared	0.095	0.116	0.26	0.261	0.232	0.246
N. observations	18,198	18,198	18,198	18,198	18,198	18,198
N. obs self-employed	1,305	1,305	1,305	1,305	1,305	1,305
Share self-employed	0.769	0.769	0.769	0.769	0.769	0.769
F-stat					932.401	915.656

***Significant to the 1% level.
NOTE: 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). Standard errors are adjusted for 109 clusters at the province of family residence.

5. Concluding remarks

This study, which is part of joint a research project between the Department of Finance of the Italian Ministry of Economy and Finance, the Universities of Milan and Insubria, and the research institution FBK-IRVAPP started two years ago, provided novel evidence on the self-employed income tax evasion in Italy. We have applied a consolidated methodology based on consumption-income comparisons between categories of taxpayers to new microdata that combines information on tax returns and consumption survey. The main results of the work can be listed as follows. First, we document that the share of self-employed income tax evasion in Italy, ranging from 30 to 40 percent of total income, is not dissimilar to that observed in different countries (United States, United Kingdom) where the same methodology has been applied. This confirm the recent view that Italy is not so exceptional internationally regarding tax evasion (D'Attoma e*t al.* (2017)). Second, we find that self-employed households located in the North of the country evade more income, about 3 percent higher, than in the rest of the country. Contrary to the popular wisdom that indicates Southern taxpayers as more evaders, we have discussed some of the possible explanations behind this result. Third, our findings point out that there are different attitudes towards tax evasion within the category of self-employed, with small entrepreneurs underreporting a lower share of income than the rest of self-employed households.

There are some policy implications that can be derived from our results. Bottom-up approaches for estimating tax evasion can be very useful instruments for complementing tax gap estimates obtained with top-down methodologies. Since two years, in Italy, in the official report on tax evasion both top-down and bottom-results regarding self-employment income tax evasion are published (MEF (2020)). In the presence of territorial differences in tax evasion behavior, as we have documented in this work, it is useful to adopt place-specific tax compliance actions in order to make the action of the tax administration more effective. Lastly, the fact that specific types of self-employed (small entrepreneurs) evade less than others highlights the importance of designing tax compliance policies, which have different costs for the administration, for particular categories of taxpayers.

References

Albarea, A., Bernasconi, M., Marenzi, A. & Rizzi, D. (2019). Income underreporting and tax evasion in Italy: Estimates and distributional effects. *Review of Income and Wealth.*

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

Bazzoli, M., Di Caro, P., Figari,F., Fiorio, C.V. & Manzo, M. (2020). Size, heterogeneity and distributional effects of self-employment income tax evasion in Italy. Dipartimento Finanze Working Paper 8/2020, available at Dipartimento Finanze—Working Papers.

Bernasconi, M. & Marenzi, A. (1997). Gli effetti redistributivi dell'evasione fiscale in Italia. Universita degli studi di Pavia.

Braiotta, A., Carfora, A., Vega Pansini, R. & Pisani, S. (2020). Tax gap and redistributive aspects across Italy. *Scienze Regionali,* 19(2), 205–226.

Cabral, A. C. G., Kotsogiannis, C. & Myles, G. (2019). Self-employment income gap in Great Britain: How much and who? *CESifo Economic Studies,* 65(1), 84–107.

Card D., Chetty R., Feldstein, M. S. & Saez, E. (2010). Expanding access to administrative data for research in the United States. Mimeo.

Carfora, A., Pansini, R. V. & Pisani, S. (2018). Regional tax evasion and audit enforcement. *Regional Studies,* 52(3), 362–373.

Clotfelter, C. T. (1983). Tax evasion and tax rates: An analysis of individual returns. *The Review of Economics and Statistics,* 363–373.

D'Attoma, J. (2019). What explains the North–South divide in Italian tax compliance? An experimental analysis. *Acta Politica,* 54(1), 104–123.

D'Attoma, J., Volintiru, C. & Steinmo, S. (2017). Willing to share? Tax compliance and gender in Europe and America. *Research & Politics,* 4(2), 2053168017707151.

Di Caro, P. (2017). The contribution of tax statistics for analysing regional income disparities in Italy. *Journal of Income Distribution,* 25(1), 1–27.

Di Caro, P. & Sacchi, A. (2020). The heterogeneous effects of labor informality on VAT revenues: Evidence on a developed country. *Journal of Macroeconomics*, 63, 103190.

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

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

Fiorio, C. V. (2011). Understanding Italian inequality trends. *Oxford Bulletin of Economics and Statistics,* 73(2):255–275.

Fiorio, C. V. & D'Amuri, F (2005). "Workers' tax evasion in Italy." *Giornale Degli Economisti e Annali di Economia:* 247–270.

González-Fernández, M. & González-Velasco, C. (2014). Shadow economy, corruption and public debt in Spain. *Journal of Policy Modeling,* 36(6), 1101–1117.

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

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

Koijen, R., Van Nieuwerburgh, S. & Vestman R. (2014). Judging the quality of survey data by comparison with "truth" as measured by administrative records: Evidence from Sweden. In: I*mproving the Measurement of Consumer Expenditures.* 308–346. University of Chicago Press.

Kukk, M., Paulus, A. & Staehr, K. (2020). Cheating in Europe: underreporting of self-employment income in comparative perspective. *International Tax and Public Finance,* 27(2), 363–390.

Lalla, M., Mantovani, D. & Frederic, P. (2019). Measurement errors and tax evasion in annual incomes: Evidence from survey data matched with fiscal data.

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

Levine, R. & Rubinstein, Y. (2018). Selection into entrepreneurship and self-employment (No. w25350). National Bureau of Economic Research.

Matsaganis, M., Benedek, D., Flevotomou, M., Lelkes, O., Mantovani, D. & Nienadowska, S. (2010). Distributional implications of income tax evasion in Greece, Hungary and Italy. Mimeo.

Ministry of Economy and Finance (2020). Relazione sull'economia non osservata e sull'evasione fiscale e contributiva. Rome.

Paulus, A. (2015). Tax evasion and measurement error: an econometric analysis of income survey data linked with tax records. ISER Working Paper 2015-10, University of Essex.

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

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. & Mróz, B. (2015). Shadow economy and tax evasion in the EU. *Journal of Money Laundering Control,* 18(1), 34–51.

Slemrod, J. & 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.

Tedds, L. M. (2010). Keeping it off the books: An empirical investigation of firms that engage in tax evasion. *Applied Economics,* 42(19), 2459–2473.

Wiseman, T. (2013). U.S. shadow economies: a state-level study. *Constitutional Political Economy,* 24(4),310–335.