# Size, characteristics and distributional effects of income tax evasion in Italy

Martina Bazzoli[1], Paolo Di Caro[2], Francesco Figari[3], Carlo V. Fiorio[4] and Marzo Manzo[2]

September 24, 2019

[1] Irvapp-FBK
[2] Ministry of Economics and Finance (MEF)
[3] University of Insubria
[4] University of Milan and Irvapp-FBK

# Introduction

- Estimating tax evasion is **daunting task**. Direct estimation is unfeasible, **indirect methodologies** is the only way to go
- Pissarides and Weber (1989, (PW)) delivered an ingenious methodology based on estimating Engel curves for food expenditures. However, **data requirements** are disappointingly high. Errors can occur because of:
  - statistical matching procedure on observables (between an income and a consumtion survey)
  - self-declared income, typically reported in income survey
- Typical outcome of expenditure-based methodologies using survey data: too much noise for eliciting any signal
- Recently, home utilities consumption (available in SILC data) was used as alternative to food consumption (Paulus, 2015; Albarea et al., 2018), but the assumption that home utilities is independent of groups is debatable and mostly untested.
- **Administrative data** provide new options in the **economists' toolbox** (e.g. Johansson (2005) for FI, Feldman and Slemrod (2007) for US, Engström and Hagen (2017) for SW)

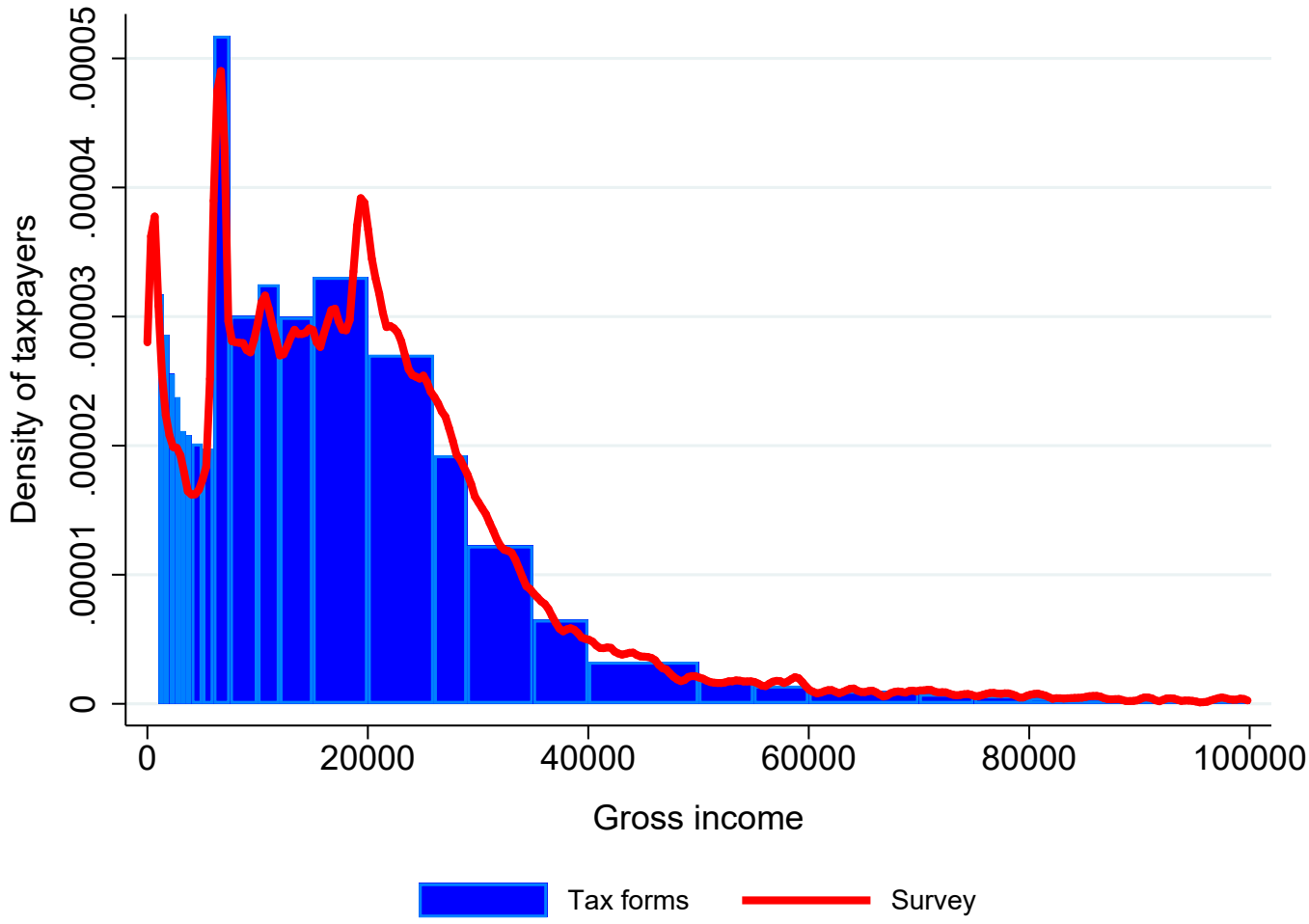
# Preview of main results

- Thanks to an **innovative collaboration** between the Italian Ministry of Economy and Finance and academic and research institutions, we have been able to provide the **first estimation of self-employment tax evasion based on the PW method for Italy**
- This model allows to estimate precise measure of permanent income as 7-year average of reported income
- We test the exclusion restriction of standard instrumental variables used for the PW methodology and found that no candidate IV variable pass the exclusion restriction test
- The richness of data allowed us to extend the model estimating heterogeneity of underreporting income. We find that tax evasion decreases with income, is larger for singles as opposed to couple and tend to be larger for college graduates.
- Tax evasion heterogeneity has strong distributional effects.
  - Underreporting taxable income goes mostly to the benefit of top decile households

# The dataset and motivation

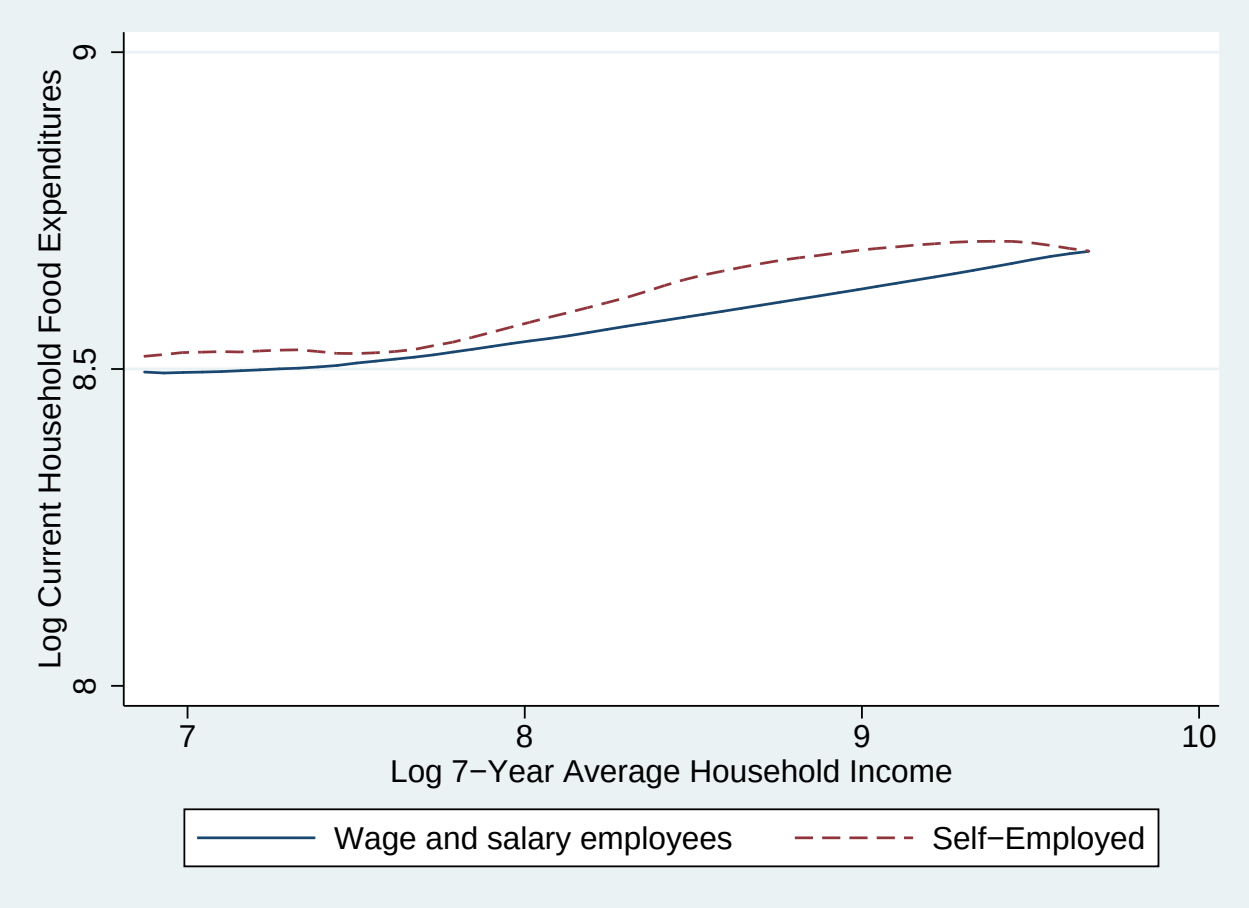
- The **estimation of tax evasion** has always been a priority in the Italian public finance academic community and it has recently acquired attention also in the international community. There is very **little sound evidence** of tax evasion estimation **in Italy**
- The MEF needs **better understanding** of the tax evasion phenomenon in Italy
- The MEF can **merge survey and administrative data** for its statutory aims
- An agreement involving MEF, Uni-Insubria, Uni-Milan and Irvapp-FBK was signed for the estimation of tax evasion with the PW methodology under condition that **data cannot exit MEF** premises
- The data was built linking the **2013 Household Budget Survey** with **2010-2016 individual tax forms**

# Sample validation




Note that the sample was not (post-)stratified by income

# Preliminary evidence




- Engel curve for self-employed (income self-reporting) households always lies above wage and salary households (third-party income reporting), confirming main assumption of PW approach
- Engel curve for wage and salary workers is roughly linear
- Heterogeneity in self-employment tax evasion is large

# The methodology I

- The reporting of expenditure on some items (food) by all groups in the population is accurate and independent on population groups identifier
- The reporting of income by some groups in the population (wage and salary workers) is accurate (third-party reported)
- The permanent income is well measured in the data (we use average of yearly income over a 7-year period)
- True (permanent) income ($y_i^{P,T}$) is larger than the reported one ($y_i^{P,R}$), where $k_{i,j}$ is larger than one if household $i$ is self-employed ($j = S$) and is equal to one otherwise ($j = E$):

$$y_i^{P,T} = k_{i,j} \cdot y_i^{P,R} \tag{1}$$

The log of $k_{i,j}$ might depend on a set of observable characteristics, $Z = z_1, z_2, ..., z_L$, which includes the constant and where $v_i$ is a zero-mean random term:

$$ln\ k_{i,j} = \sum_{l=1}^{L} \mu_l z_l + v_i. \tag{2}$$

# The methodology II

Using an indicator $S_i$ that takes values equal one if the household $i$ is self-employed, and zero otherwise, we can write Engel curve as:

$$ln\, c_i = \beta \cdot ln\, y_i^{P,R} + \sum_{l=1}^{L} \underbrace{S_i \cdot \beta \cdot \mu_l}_{\gamma_l} \cdot z_l + X_i' \alpha + \epsilon_i + S_i \cdot \beta \cdot v_i \qquad (3)$$

The proportion of true income that is reported by self-employed households is:

$$\widehat{\kappa}_{iS} = exp\left[ - \frac{\sum_{l=0}^{L} \widehat{\gamma}_l \cdot E\left(z_l\right)}{\widehat{\beta}} \right] \qquad (4)$$

and $(1 - \widehat{\kappa}_{iS})$ is the proportion of unreported income.

# The sample

- Focus of the analysis is the **whole sample**, hence no sample selection is performed
- Observation are at the household level (because HBS data refer to the household)
- We identify self-employed families based on prevalent income in the household (self-employed vs wage and salary income)

# Some reflections before estimation

- In surveys current income is typically the only candidate measure of $y^P$. To deal with endogeneity an IV strategy is often implementing using as instruments, education, capital income or home characteristics.
  - Is this meaningful?
- Similarly to Engström and Hagen (2017), we use a 7-year average (from $t-3$ to $t+3$) income measure of reported income, which is a good measure of permanent income
- Similarly to Hurst et al. (2014), we use both Pre- and Post-Tax Total Household Income.
  - Here we show only Pre-Tax results (robust results using Post-Tax variable)

# Estimation: OLS

Table: Engel curves for (log) food expenditures and alternative measures of income. A self-employed household has at least 50% of its income from self-employment.

| | (A) | (B) | (C) | (D) | (E) | (F) |
|---|---|---|---|---|---|---|
| | Income = Pre-tax Total Family Income | | | | | |
| Self-employed | 0.183*** | 0.184*** | 0.064*** | 0.072*** | 0.050*** | 0.060*** |
| | (0.017) | (0.017) | (0.016) | (0.017) | (0.016) | (0.017) |
| Current inc. | 0.195*** | | 0.106*** | | 0.083*** | |
| | (0.009) | | (0.007) | | (0.008) | |
| Perm. inc. (over 7 yrs) | | 0.226*** | | 0.128*** | | 0.104*** |
| | | (0.008) | | (0.008) | | (0.009) |
| $(1-\kappa)$ | 0.609*** | 0.556*** | 0.450*** | 0.431*** | 0.451*** | 0.439*** |
| | (0.038) | (0.035) | (0.085) | (0.075) | (0.105) | (0.088) |
| Controls | No | No | Yes | Yes | Yes | Yes |
| Additional controls | No | No | No | No | Yes | Yes |
| R-squared | 0.095 | 0.113 | 0.203 | 0.206 | 0.21 | 0.211 |
| N. obs | 18220 | 18305 | 18220 | 18305 | 18220 | 18305 |
| N. obs Self-Employed | 1775 | 1807 | 1775 | 1807 | 1775 | 1807 |
| Share Self-Employed | 0.866 | 0.859 | 0.866 | 0.859 | 0.866 | 0.859 |

Notes: 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. Additional controls include also household head education and building property wealth (cadastral values). Standar errors are adjusted for 109 clusters at the province of family residence.

# Testing the exclusion restriction of IV estimates

- Typically, application of expenditure-based methods using current income involves the use of IV for instrumenting current income, as a good measure of permanent income is unavailable in surveys.
- Here a good measure of permanent income is available (average of yearly income over 7-year period)
- Good instruments should be relevant (first-stage significant) and uncorrelated with the error term (exclusion restriction)
- For testing the exclusion restriction we estimate a reduced form regression of the Engel curve, including potential instruments: if statistically significant, the exclusion restriction is not satisfied

# IV estimation: testing the relevance of instruments

Table: Engel curves for (log) food expenditure using permanent (7-year average) income and IV estimation using alternative instruments. A self-employed household has at least 50% of its income from self-employment.

| | (A) | (B) | (C) |
|---|---|---|---|
| | Income = Pre-tax Total Family Income | | |
| Self-employed | 0.114*** | 0.087*** | 0.095*** |
| | (0.021) | (0.017) | (0.017) |
| Perm. inc. (over 7 yrs) | 0.292*** | 0.195*** | 0.228*** |
| | (0.031) | (0.024) | (0.012) |
| $(1-\kappa)$ | 0.322*** | 0.359*** | 0.340*** |
| | (0.041) | (0.058) | (0.050) |
| R-squared | 0.167 | 0.201 | 0.191 |
| N. obs | 18305 | 18305 | 18305 |
| N. obs Self-Employed | 1807 | 1807 | 1807 |
| Share Self-Employed | 0.859 | 0.859 | 0.859 |
| F-stat | 213.264 | 908.649 | 387.433 |

Notes: Controls include household head gender and age, 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 province of residence dummies.
Column (A) building property wealth (cadastral values) used as control and household head education as instrumental variable. Column (B) includes household head education as controls and property wealth (cadastral values) as instrumental variable. Column (C) includes household head education, building property wealth (cadastral values) as instrumental variables.
Standar errors are adjusted for 109 clusters at the province of family residence.

# IV estimation: testing the exclusion restrictions

Table: Testing the exclusion restrictions in Engel curves for (log) food expenditure using permanent (7-year average) income. A self-employed household has at least 50% of its income from self-employment.

| | (A) Income = Pre-tax Total Family Income | (B) Income = Post-tax Total Family Income |
|---|---|---|
| Self-employed | 0.060*** | 0.061*** |
| | (0.017) | (0.017) |
| Perm. inc. (over 7 yrs) | 0.104*** | 0.109*** |
| | (0.009) | (0.010) |
| Test exclusion: household head education | 0.000 | 0.000 |
| Test exclusion: property wealth (cadastral values) | 0.000 | 0.000 |
| Test exclusion: all potential instruments jointly | 0.000 | 0.000 |
| R-squared | 0.211 | 0.210 |
| N. obs | 18305 | 18302 |

Notes: Controls include household head gender and age, education (primary, secondary or tertiary) of the spouse interacted with in-couple dummy, household size, a dummy for presence of kids, family consumption of sin goods, a full set of province of residence dummies.
The exclusion restriction test shows the p-value of an F-test of the significance of candidate instruments in the reduced form regression.

# Estimating heterogeneity in the share of unreported income

Table: Estimates of $(1 - k_{ij})$, using eq. (2). A self-employed household has at least 50% of its income from self-employment.

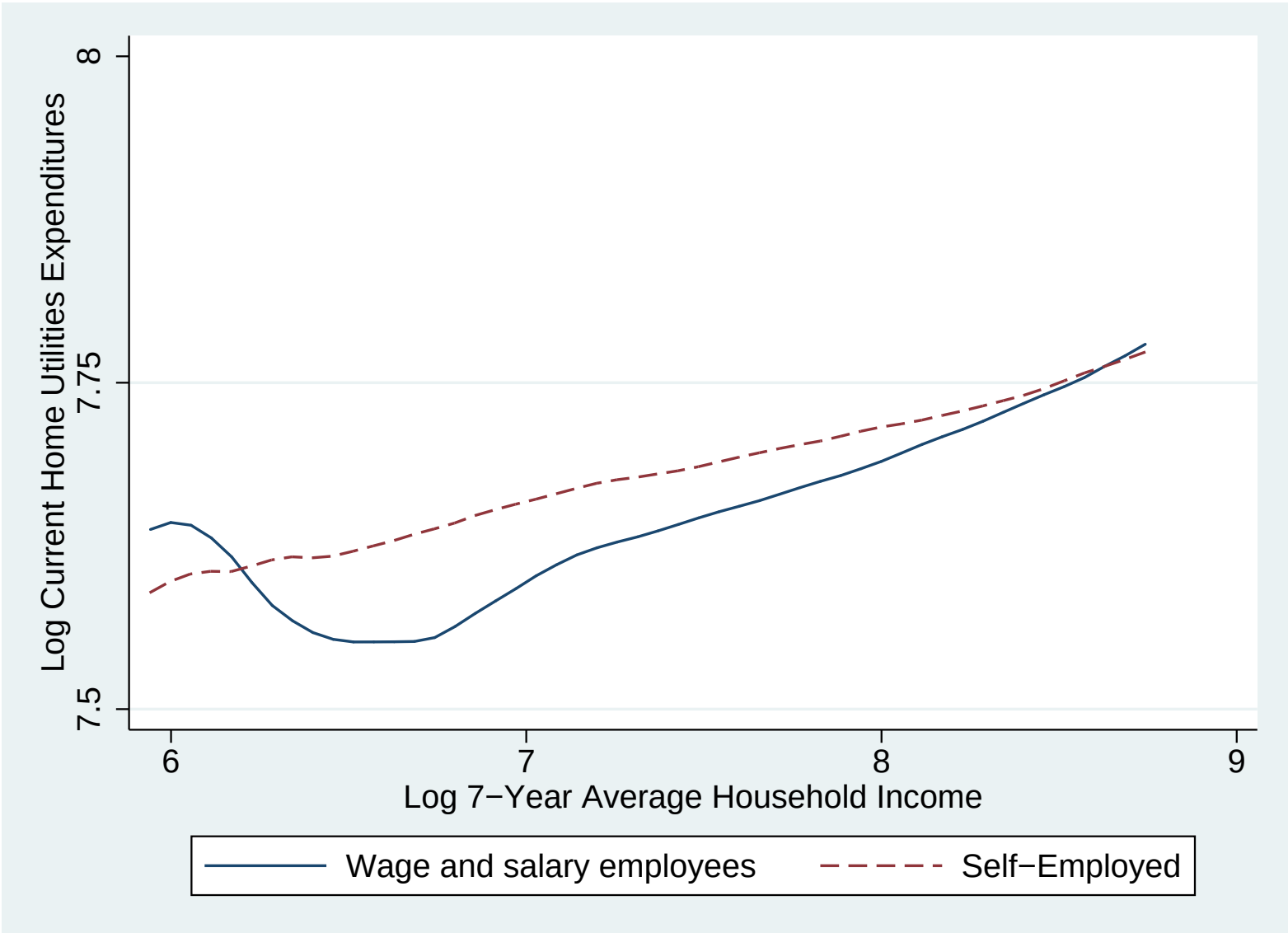
| | (A) | (B) | (C) | (D) | (E) |
|---|---|---|---|---|---|
| | Income = Pre-tax Total Family Income | | | | |
| Self-empl. | 0.060*** | 0.580*** | 0.536*** | 0.575*** | 0.598*** |
| | (0.017) | (0.159) | (0.156) | (0.153) | (0.154) |
| Perm. inc. (over 7 yrs) | 0.104*** | 0.115*** | 0.113*** | 0.114*** | 0.115*** |
| | (0.009) | (0.010) | (0.010) | (0.010) | (0.010) |
| Self-empl. × Perm. inc. (over 7 yrs) | | -0.052*** | -0.041*** | -0.046*** | -0.049*** |
| | | (0.015) | (0.015) | (0.015) | (0.015) |
| Self-empl. × Couple | | | -0.101*** | -0.093*** | -0.094*** |
| | | | (0.032) | (0.033) | (0.033) |
| Self-empl. × Head College educ. | | | | 0.070** | 0.010 |
| | | | | (0.034) | (0.044) |
| Self-empl. × Head College educ. × North | | | | | 0.105* |
| | | | | | (0.058) |
| R-squared | 0.21 | 0.211 | 0.211 | 0.211 | 0.211 |
| N. obs | 18302 | 18302 | 18302 | 18302 | 18302 |
| N. obs Self-empl. | 1805 | 1805 | 1805 | 1805 | 1805 |
| Share Self-empl. | 0.859 | 0.859 | 0.859 | 0.859 | 0.859 |

Notes: 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. Additional controls include also household head education and building property wealth (cadastral values).
Standar errors are adjusted for 109 clusters at the province of family residence.

# Robustness checks: Alternative dependent variable

- When food consumption is unavailable can we still use PW? Paulus (2015), exploiting SILC, propose to use yearly home utility expenditures. Similarly did Albarea et al. (2018).
  - Can we **test** whether we get conflicting results using home utilities expenditures instead of food expenditure?
  - What is the analogue of descriptive figure above, when home utilities instead of food expenditure is used?

# Robustness checks: Alternative dependent variable



# Robustness checks: Alternative dependent variable

Table: Engel curves for (log) home utility expenditures and alternative measures of income. A self-employed household has at least 50% of its income from self-employment

| | (A) | (B) | (C) | (D) | (E) | (F) |
|---|---|---|---|---|---|---|
| | Income = Pre-tax Total Family Income | | | | | |
| Self-employed | 0.131*** | 0.126*** | 0.076*** | 0.079*** | 0.049** | 0.053** |
| | (0.022) | (0.022) | (0.020) | (0.020) | (0.021) | (0.021) |
| Current inc. | 0.208*** | | 0.116*** | | 0.077*** | |
| | (0.012) | | (0.009) | | (0.009) | |
| Perm. inc. (over 7 yrs) | | 0.240*** | | 0.139*** | | 0.096*** |
| | | (0.012) | | (0.009) | | (0.010) |
| $(1-k)$ | 0.467*** | 0.409*** | 0.480*** | 0.432*** | 0.474*** | 0.428*** |
| | (0.056) | (0.052) | (0.087) | (0.079) | (0.133) | (0.118) |
| Controls | No | No | Yes | Yes | Yes | Yes |
| Additional controls | No | No | No | No | Yes | Yes |
| R-squared | 0.102 | 0.119 | 0.195 | 0.199 | 0.212 | 0.214 |
| N. obs | 18214 | 18299 | 18214 | 18299 | 18214 | 18299 |
| N. obs Self-Employed | 1774 | 1806 | 1774 | 1806 | 1774 | 1806 |
| Share Self-Employed | 0.866 | 0.860 | 0.866 | 0.860 | 0.866 | 0.860 |

Notes: 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. Additional controls include also household head education and building property wealth (cadastral values). Standar errors are adjusted for 109 clusters at the province of family residence.

# Robustness checks: Alternative definition of self-employed households

Table: Engel curves for (log) food expenditures and alternative measures of income. A self-employed household has at least 25% of its income from self-employment

| | (A) | (B) | (C) | (D) | (E) | (F) |
|---|---|---|---|---|---|---|
| | Income = Pre-tax Total Family Income | | | | | |
| Self-employed | 0.173*** | 0.171*** | 0.051*** | 0.058*** | 0.038** | 0.047*** |
| | (0.016) | (0.015) | (0.017) | (0.016) | (0.016) | (0.016) |
| Current inc. | 0.192*** | | 0.106*** | | 0.082*** | |
| | (0.009) | | (0.007) | | (0.008) | |
| Perm. inc. (over 7 yrs) | | 0.224*** | | 0.127*** | | 0.103*** |
| | | (0.008) | | (0.008) | | (0.009) |
| $(1-\kappa)$ | 0.594*** | 0.534*** | 0.385*** | 0.366*** | 0.371*** | 0.364*** |
| | (0.039) | (0.036) | (0.097) | (0.081) | (0.120) | (0.095) |
| Controls | No | No | Yes | Yes | Yes | Yes |
| Additional controls | No | No | No | No | Yes | Yes |
| R-squared | 0.097 | 0.114 | 0.203 | 0.206 | 0.21 | 0.211 |
| N. obs | 18220 | 18305 | 18220 | 18305 | 18220 | 18305 |
| N. obs Self-Employed | 2374 | 2405 | 2374 | 2405 | 2374 | 2405 |
| Share Self-Employed | 0.743 | 0.739 | 0.743 | 0.739 | 0.743 | 0.739 |

Notes: 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. Additional controls include also household head education and building property wealth (cadastral values). Standar errors are adjusted for 109 clusters at the province of family residence.

# What are the budgetary and distributional effects of tax evasion?

- EUROMOD used to calculate social insurance contributions, taxes and benefits, disposable income
  - Self employment income derived according to the estimated shares of unreported income
  - Baseline: no full tax compliance
  - Scenario: full tax compliance
- Input data Sample:
  - HBS 2013
  - Income variables: admin data 2013 (updated to 2018)
- Policy system
  - 2018
- Distributional analysis performed looking at deciles of household-equivalent disposable income

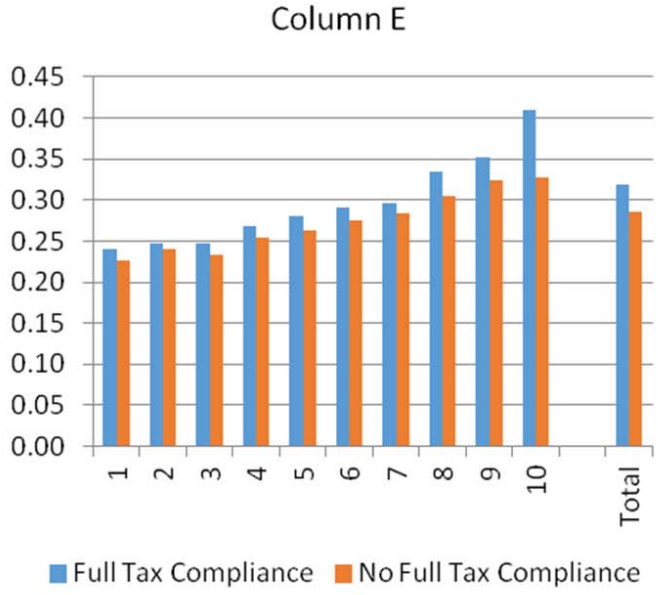
# Budgetary effects

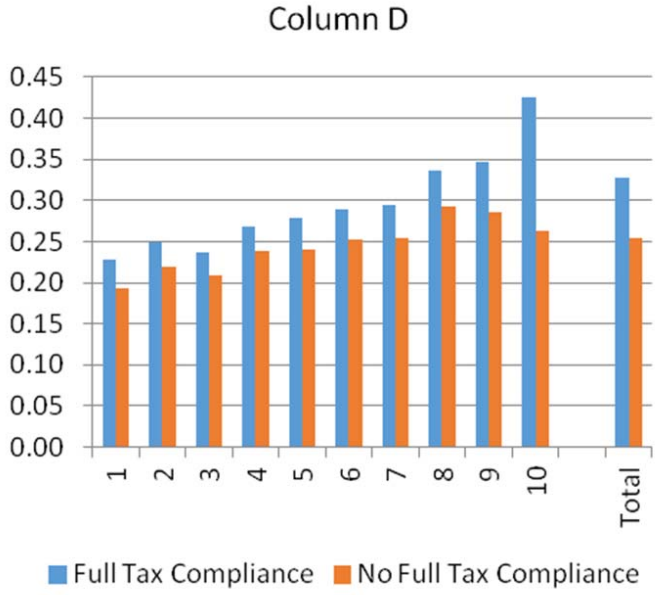
| Budgetary effects | | E | D | C | B |
|---|---|---|---|---|---|
| | | | Changes in €million, annual | | |
| Self employment income | 121,991.71 | 48,962.29 | 95,365.82 | 82,637.61 | 74,907.62 |
| Taxable income | 793,552.80 | | | | |
| IRPEF | 179,049.91 | 19,942.76 | 37,258.15 | 31,245.73 | 26,968.90 |
| SE SSCs | 25,078.99 | 2,999.42 | 8,214.25 | 8,574.33 | 9,504.25 |
| Transfers | 256,998.23 | 26.31 | 70.41 | 74.51 | 109.62 |
| | | | Changes in % | | |
| IRPEF | | 11.14 | 20.81 | 17.45 | 15.06 |
| SE SSCs | | 11.96 | 32.75 | 34.19 | 37.90 |
| Transfers | | -0.01 | -0.03 | -0.03 | -0.04 |
| Disposable income | | -3.46 | -6.41 | -5.72 | -5.30 |

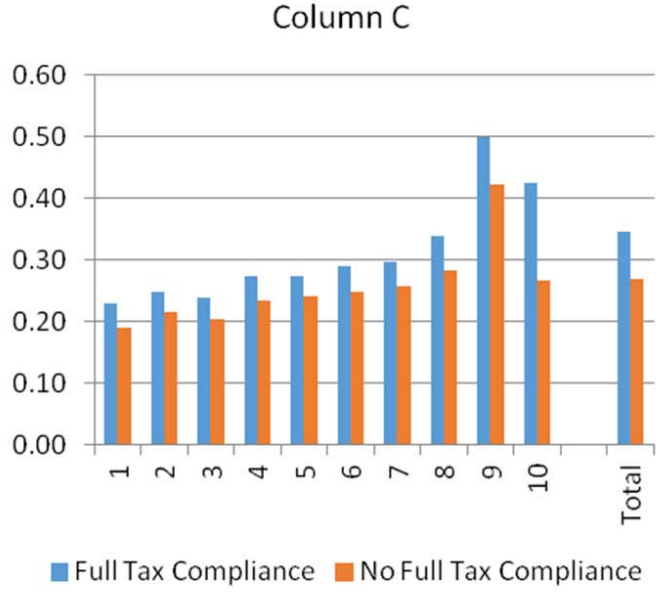
# Redistributive effects

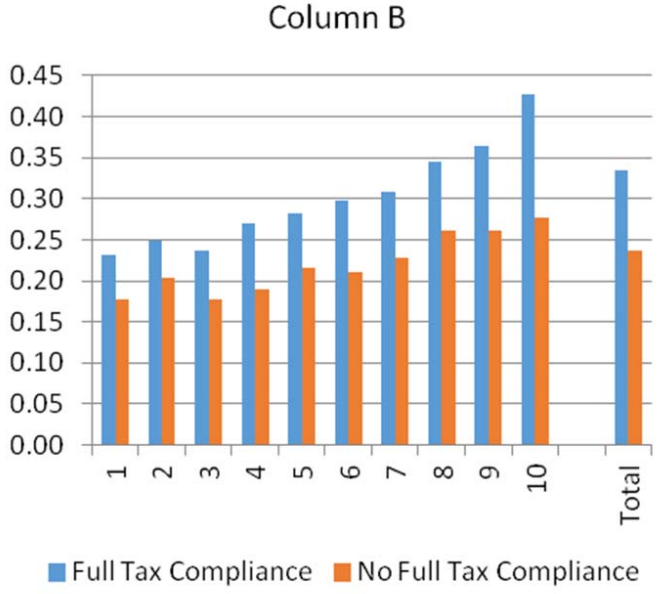
| | No Full tax compliance | Full tax compliance |
|---|---|---|
| | *Scenario Table 6, column E* | |
| Taxable income, Gini | 0.430 | 0.458 |
| Disposable income, Gini | 0.398 | 0.378 |
| *RE* | *0.032* | *0.081* |
| | *Scenario Table 6, column D* | |
| Taxable income, Gini | 0.430 | 0.468 |
| Disposable income, Gini | 0.411 | 0.383 |
| *RE* | *0.019* | *0.084* |
| | *Scenario Table 6, column C* | |
| Taxable income, Gini | 0.430 | 0.460 |
| Disposable income, Gini | 0.400 | 0.377 |
| *RE* | *0.029* | *0.083* |
| | *Scenario Table 6, column B* | |
| Taxable income, Gini | 0.430 | 0.454 |
| Disposable income, Gini | 0.391 | 0.371 |
| *RE* | *0.039* | *0.083* |

# Changes in Average Tax Rates

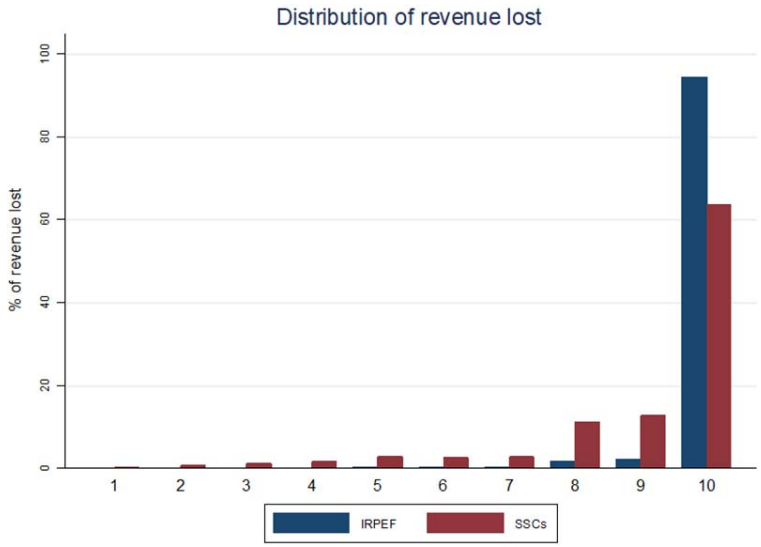


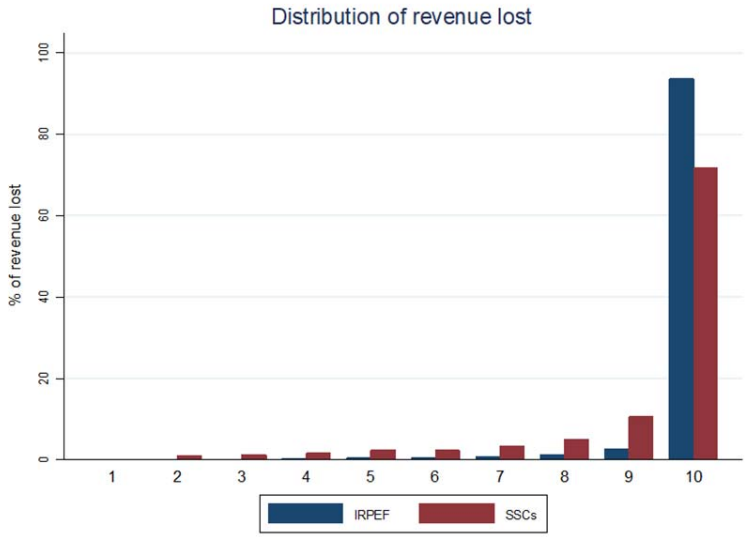



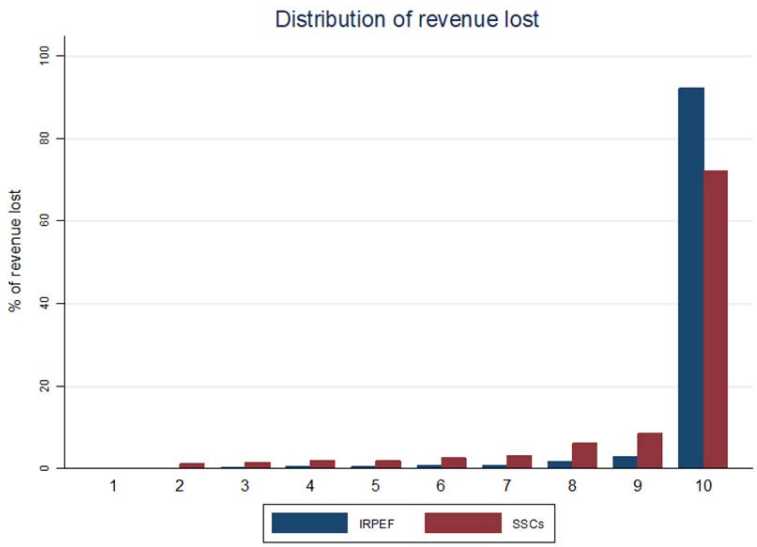



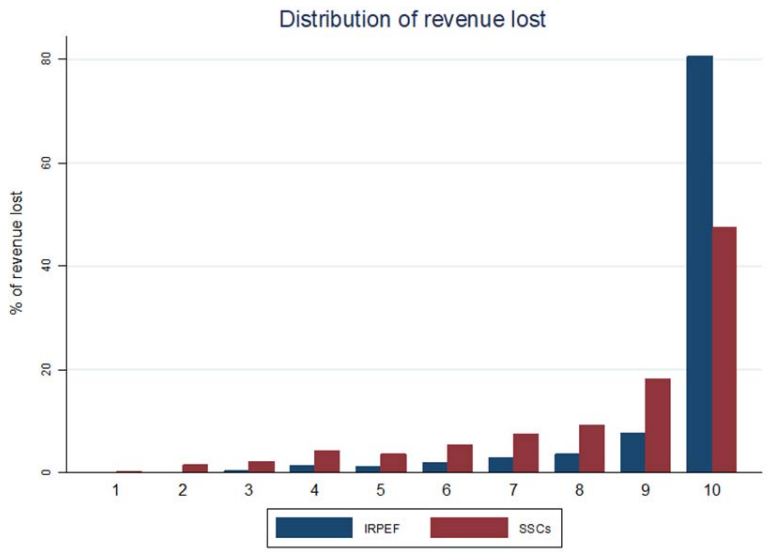



# Where does the lost revenue go?

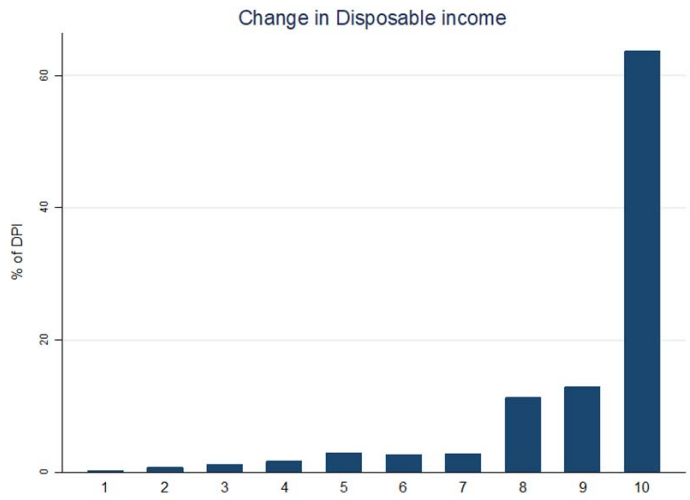


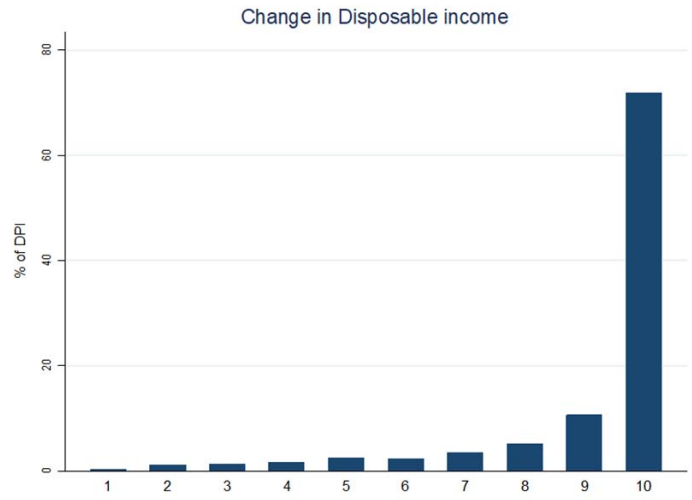



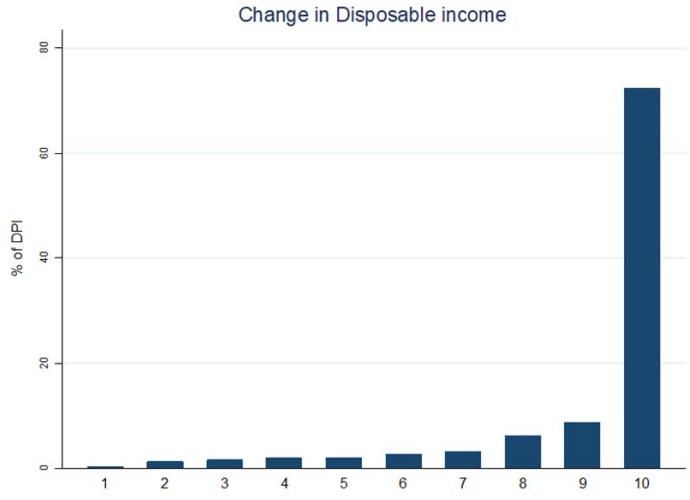



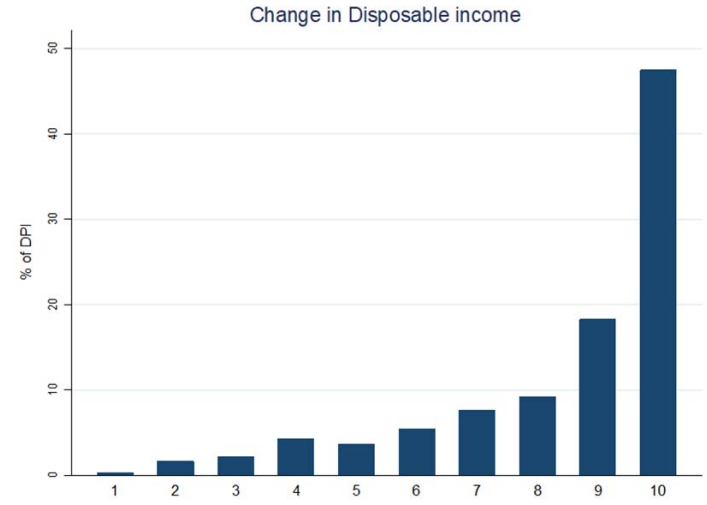



# How does the disposable income change?



# Conclusions and future developments

- First time a PW approach is performed for Italy with statistically significant results.
  - These Italian data are unique. Only comparable alternative: Engström and Hagen (2017)
- Italy is confirmed as a high tax evasion country (**self-employment underreport over 40%** of their Pre- and Post-Tax income as opposed to wage and salary income).
- Results are robust to alternative definitions of dependent variable and permanent income
- Redistributive effects are large and important
  - tax evasion benefits mostly the households in top deciles
- Limitations:
  - we cannot disentangle civil servants from private sector employees (Paulus (2015))
  - we currently have no information on capital income & wealth

# Main references


Albarea, A., Bernasconi, M., Di Novi, C., Marenzi, A., Rizzi, D., and Zantomio, F. (2018). Income under reporting and tax evasion in italy: estimates and distributive effects. Rapporto di Valutazione n. 8, 2015:24, Senato della Repubblica. Ufficio valutazione impatto. Senato della Repubblica.

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

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

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.

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

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.