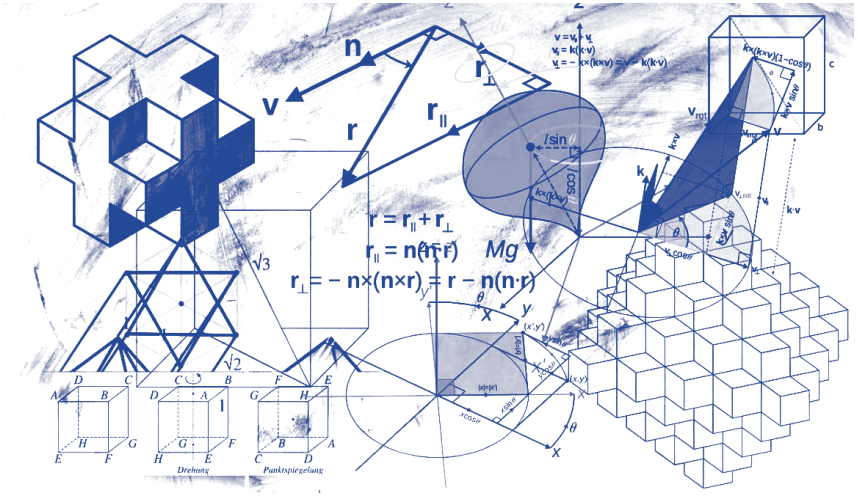
SECOND INTERNATIONAL WORKSHOP - MOSPI PROJECT

The Treasury DYnamic Microsimulation Model (T-DYMM): structure, preliminary results and future implementations

PANEL 1

T-DYMM 3.0: Data, Model Structure and Recent Innovations

Presenters: Riccardo Conti, Stefano Boscolo
Discussants: Gijs Dekkers, Maria Luisa Maitino



With financial support from the European Union – EaSI Progress Program

Outline

- History and general features
- Data
- General structure of modules
- Demographic Module

- Focus 1: income correction for the self-employed
- Focus 2: weight calibration

General features of T-DYMM

- It is a Dynamic Microsimulation Model (medium and long-term simulations)
- It treats time as discrete
- It has a sequential structure
- Socio-economic events occur according to conditional transition probabilities (estimated externally)
- It uses alignment procedures on demographic and macroeconomic dimensions
- It runs on the Liam 2.0 platform, developed by the Federal Planning Bureau (Belgium), with testing and funding by CEPS/INSTEAD (Luxembourg) and IGSS (Luxembourg) (liam2.plan.be)

Development of T-DYMM

- T-DYMM has been developed in 3 phases:
  1. 1° European Project (**2010-2012**): based on MIDAS-IT (derived from MIDAS-BE) and EconLav, developed in Liam 1.0
  2. 2° European Project (IESS, **2014-2016**): new and improved data, move to Liam 2.0, update of legislation, addition of private pension module, unemployment benefits
  3. 3° European Project (MOSPI, **2019-2021**): new and improved data, improvement of sample representativeness, inclusion of working pensioners, expansion of Disability Module, development of a Tax-Benefit Module, a Wealth Module and a Migration Module

Areas of focus for T-DYMM (1)

- T-DYMM has generally been used to assess the *adequacy* of the Italian pension system. Published results inculde:
  - Average retirement age
  - Average duration of retirement at death
  - Replacement rate at retirement
  - Gini index
  - Income quintile share ratio (S80/S20)
  - At-Risk-of-Poverty Rate (AROP)
- Results have generally been proposed on a number of sensitivity and policy scenarios

Areas of focus for T-DYMM (2)

- T-DYMM 1.0: adequacy of public pensions
- T-DYMM 2.0: adequacy of public and private pensions and of unemployment benefits
- T-DYMM 3.0: adequacy of pensions, unemployment benefits and tax-benefit system

Looking ahead

- Upcoming deliveries:
  - MOSPI project (final conference in December 2021)
  - Migration module methodology discussed in Note delivered by Special Interest Group on Microsimulation (Ingrid-2 Project)
  - T-DYMM 3.0 baseline results to be included in Note in dissemination with 2021 Pension Adequacy Report, an update to 'What are the consequences of the AWG 2018 projections and hypotheses on pension adequacy?'
  - Presentation paper and papers on policy scenarios

The AD-SILC dataset – main structure (1)

- The core of T-DYMM's dataset is composed by matching:
  - **Survey data** contained in the European Union Statistics on Income and Living Conditions (EU-SILC), delivered for Italy by the Italian National Institute of Statistics (ISTAT)
  - **Administrative data** from the Italian National Institute of Social Security (INPS)

- The merging procedure is conducted through individual tax codes (*codici fiscali*) that are subsequently anonymized. We call the merged dataset **AD-SILC**

The AD-SILC dataset – main structure (2)

- AD-SILC is an **unbalanced panel** dataset that comprises all SILC waves from 2004 to 2017 and all information contained in INPS concerning surveyees
- Information content:
  - **SILC**: longitudinal data on socio-economic characteristics (up to 4 years), a total of **254,212 individuals**
  - **INPS**: longitudinal data on pensions (disability, old-age, survivor) and working history (occupational status, income evolution, contribution accrual), a total of **6,182,926 observations** over the 1922-2018 period

The AD-SILC dataset – 3.0 innovations

- Addition of **five SILC waves** (2013-2017)
- Merge of information from **Tax Returns** and **Cadaster** for the 2010, 2012, 2014 and 2016 corresponding SILC waves
- Statistical matching to include information from the Survey on Household Income and Wealth (**SHIW**) conducted by the Bank of Italy

AD-SILC uses: Analyses, regressions and simulations

- Analyses of dynamics within the labor market
- Regression parameters used in T-DYMM are based on estimations run on the **entire** AD-SILC dataset and are used to model a number of processes
- Simulations are based on a **single extract** of AD-SILC (SILC 2016)

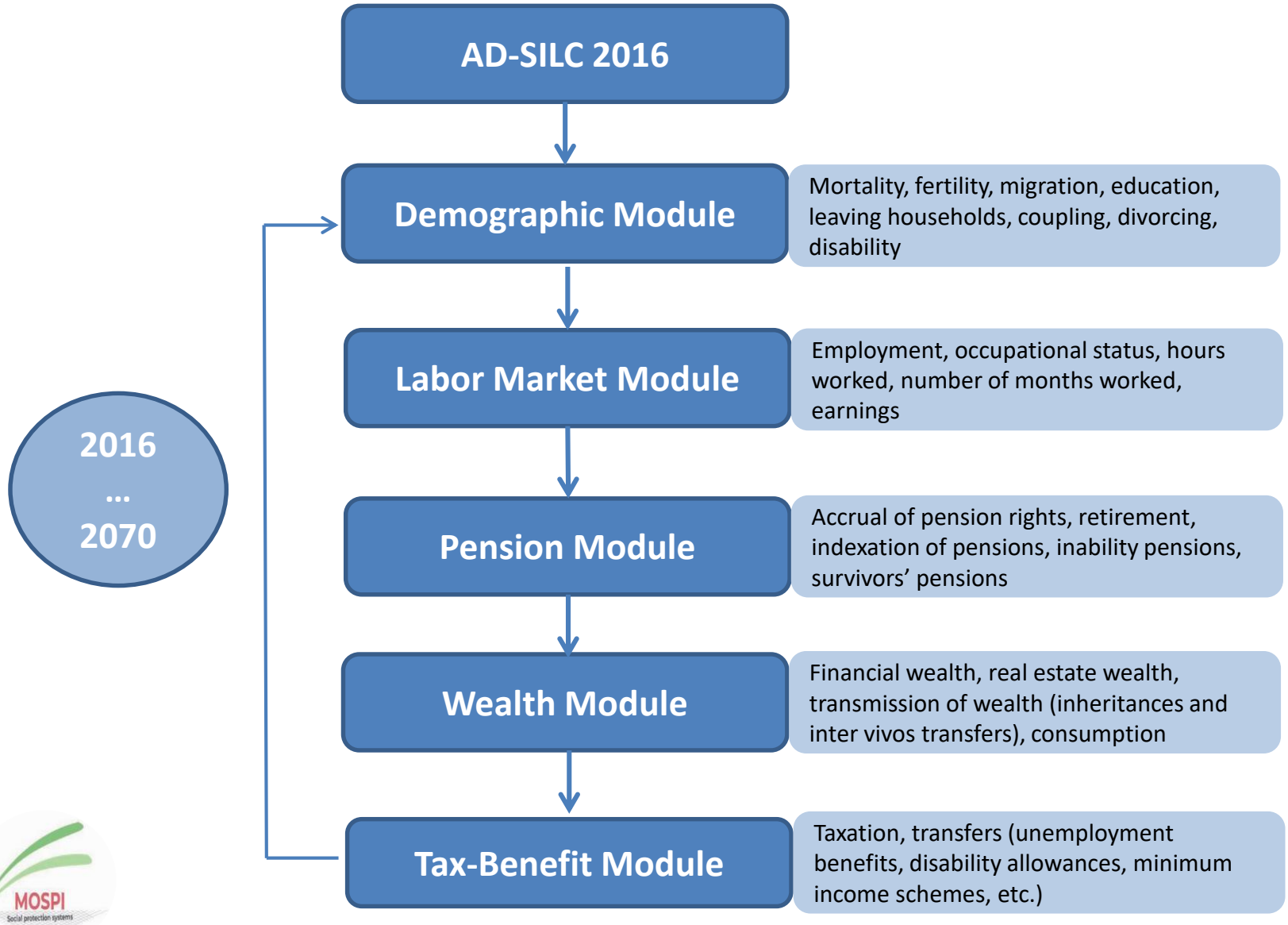
Macro data and alignments

- Exogenous data are used to align a number of patterns within the simulations:
  - **Europop projections**: mortality rate, fertility rate, immigration and emigration by gender
  - **Ageing Report** assumptions: employment rate, inflation, GDP, productivity, disability rate, returns on risk-free assets
  - **Italian Finance Department**: number of households paying rents, total beneficiaries of specific tax expenditures and substitute regimes
  - **ISTAT**: leaving household of origin, age and country of birth of migrants, education, acquisitions of houses, propensity to consume
  - **INPS**: occurrence of disability allowances and inability pensions
  - **COVIP** (Italian Vigilance Committee on Private Pension Plans): enrollment in private pension plans

Future implementations

- Information from Registers (AIRE/ANPR) to gain understanding of the migration phenomenon
- Extend the link with Tax Returns and Cadaster data
- Add information related to health (*SDO*, *Tessera sanitaria*)

T-DYMM 3.0 Module structure



The Demographic Module - processes

1. Ageing and mortality (no heterogeneity)*
2. Births (probabilistic)*
3. International migration (cloning procedure)*
4. Disability (probabilistic, aligned)*
5. Education (dependent on parents' education)**
6. Leaving household (deterministic)**
7. Coupling / marriage and divorce (probabilistic)***

Aligned to:
* Europop/AWG Projections
** ISTAT
*** Ad-hoc alignments

The Migration sub-module (1)

Why do we need a Migration sub-module?

Text of the MOSPI project

*"To duly keep into account the relevance of the phenomenon of migration both for the **demography** and for the **labor market** of Italy, a "migration module" has to be implemented. At present time, T-DYMM works as a closed model, thus neither considering immigration nor emigration*

…

*A number of **sensitivity** scenarios will be built. The role of demographic projections is crucial for the prospected sustainability of social protection systems in the medium-long term. For this reason, we will focus on the effect of changing the underlying components, in particular stressing fertility and migration."*

- Demography (sample bound to shrink without immigrants)
- Labor Market (migrant workers are expected to behave differently)

The Migration sub-module (2)

Common issues with data on migrants

- Registers do not account for the whole phoenomenon of migration
- Survey data lack representativeness
- Definition of immigrants: foreign-born vs non-citizens
- General lack of data (especially for emigrants)

The Migration sub-module (3)

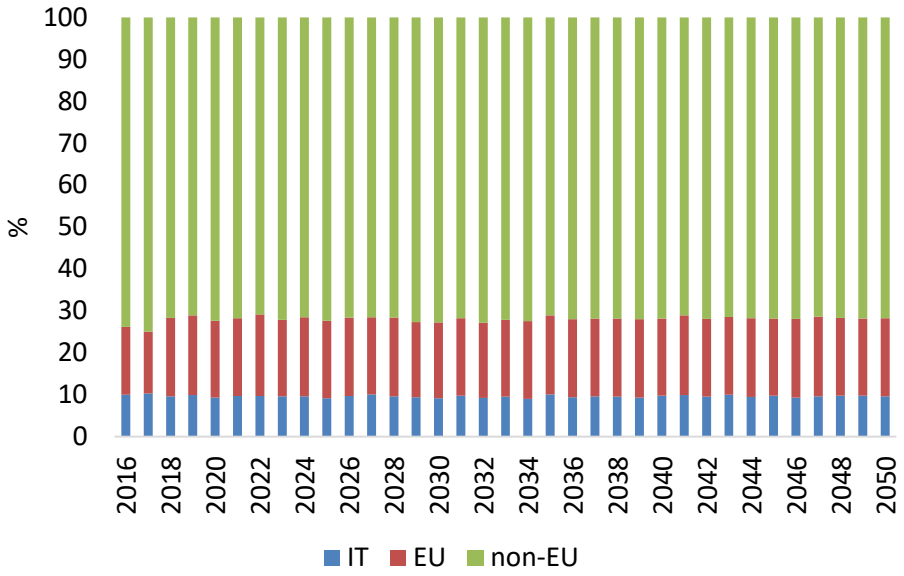
Our proposed methodology

- We simulate immigration and emigration separately
- We follow Dekkers (2015) and Chénard (2000) and implement a 'cloning procedure' for households using Chénard's Pageant algorithm
- We focus on three essential dimensions to define migrants: age, gender and area of birth (IT, EU, non-EU)
- Immigrant households derive household composition from cloned households. Education achievements are attributed individually
- Inflows and outflows of migrants are aligned to Europop projections, education (for immigrants) and area of birth is assumed constant (by age group) according to respectively OECD and ISTAT data
- Immigrants 'start fresh'
- Emigrants are deleted from the simulation (overestimation problem)

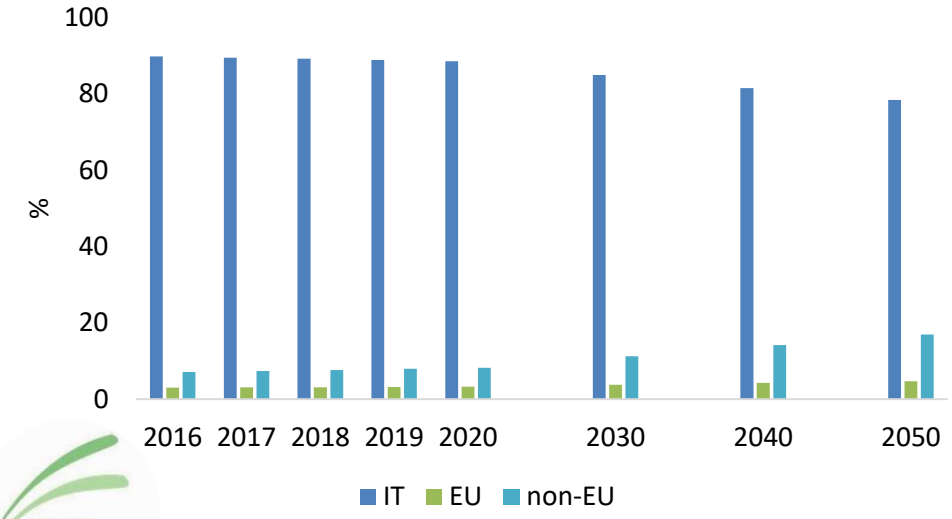
Sample structure through the simulation period - area of birth

- Following recent trends, future inflows are mostly of non-EU origin

Immigration flows by area of birth - 2016-2050



Resident population by area of birth, 2016-2050

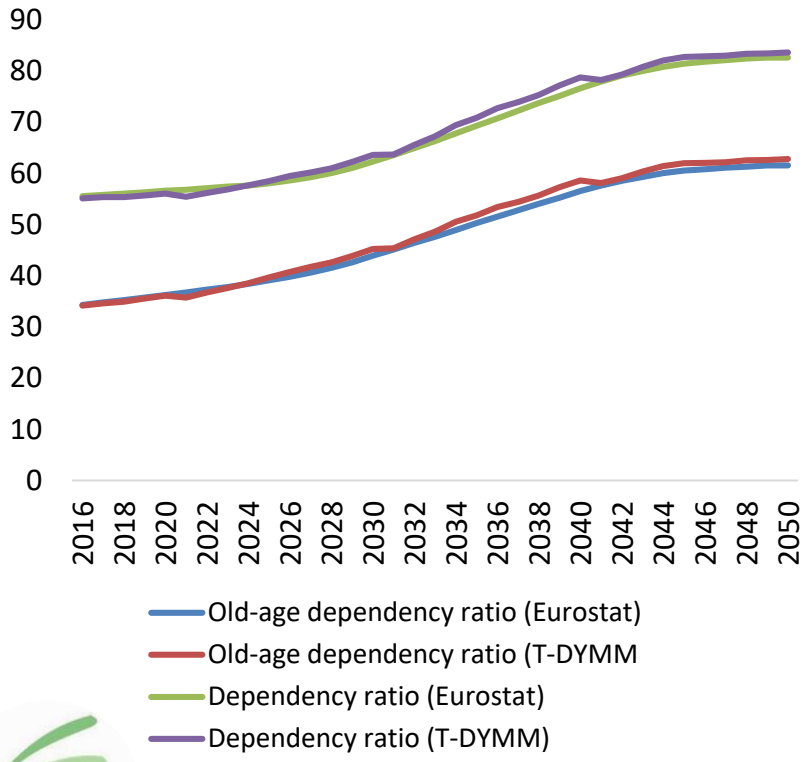


- As a result of the projections on mortality and fertility rates and of migration flows, the resident population goes from 89% Italian-born (2019) to 78% (2050)

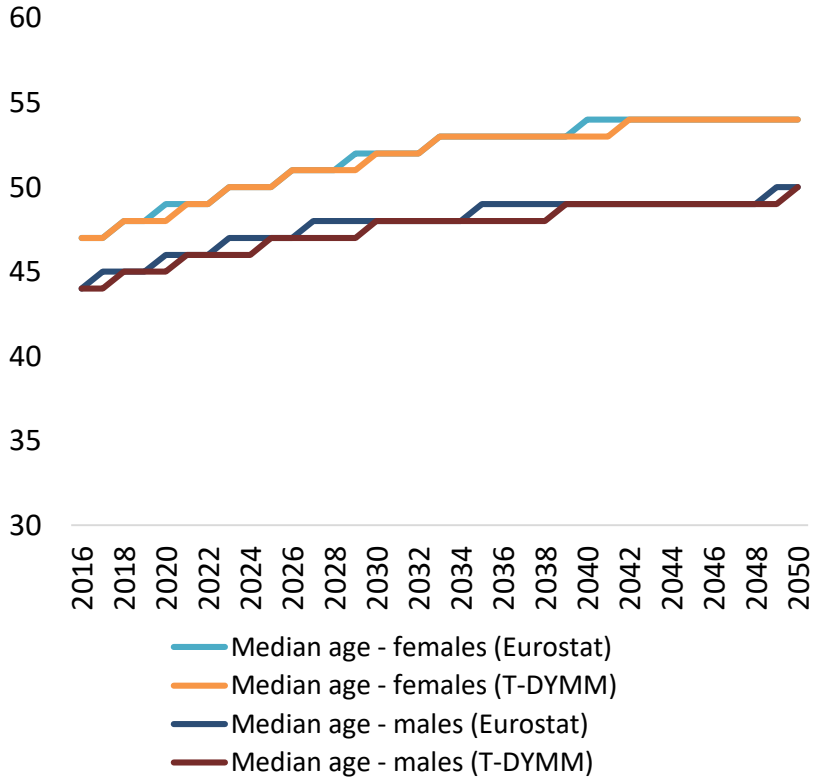
Age structure through the simulation period

- Despite the contribution of immigrants, dependency ratios and median age increase significantly over the simulation period. T-DYMM results are (by construct) in line with Eurostat projections

Dependency Ratios 2016-2020



Median age by gender - 2016-2050



Future implementations

- Heterogeneity in mortality
- Simulation of return migration
- Development of Health Module

Self-employment income correction

- Background:
  - Estimates for the Labor Market Module are based on AD-SILC 2004-2017
  - For specific categories of self-employed workers, INPS archives collect earnings that do not match tax returns data
  - Earnings as collected in INPS archives are constant below specific income thresholds and are used to compute social insurance contributions (SICs)

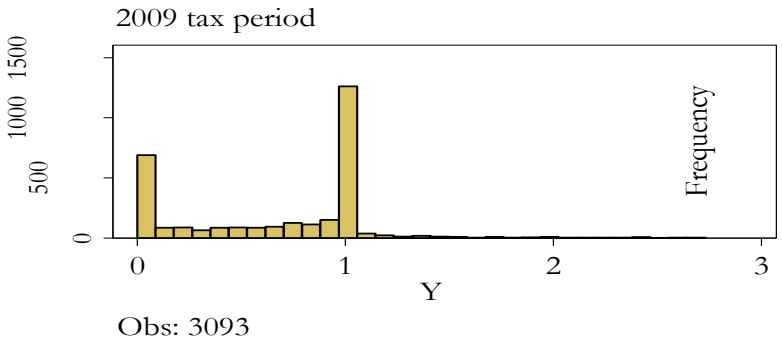
- Who are these self-employed workers?
  - Craftsmen and traders with earnings lower than an income threshold *T1*, which depends on the number of months worked (e.g. for those who worked the whole year, the threshold was set to 15,548 euros in 2015)
  - Farmers with earnings lower than an income threshold *T2*, which depends on the number of days of contribution accrued in a year (e.g. for those who accrued 312 days of contribution, the threshold was set to 17,175.6 euros in 2015)

- Example: one earns less than *T1* and thus will declare *T1-k* for tax purposes (where $k>0$ and $k<T1$), but *T1* is the income value that will be reported in INPS archives

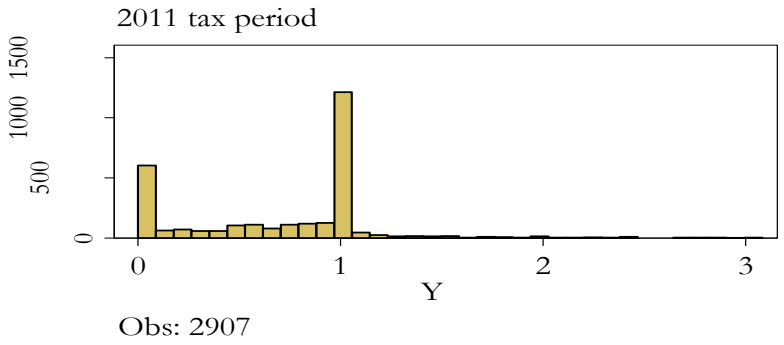
Descriptive statistics (1)

- The graph shows the ratio between earnings as declared for tax purposes and earnings as collected in INPS archives (Y) for workers i) with INPS earnings lower than 70,115 (71,737; 75,883; 76,872) euros for the 2009 (2011; 2013; 2015) tax period; ii) exclusively self-employed (craftsmen, traders, farmers or freelancers); iii) aged 16-79 and not in education.

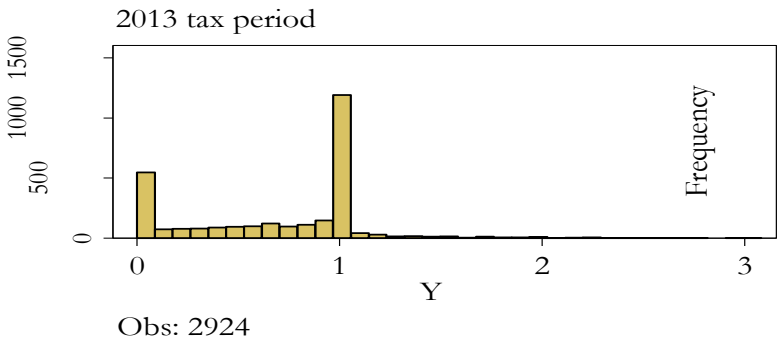
- We found that roughly 50% of observations have INPS earnings equal to declared earnings for tax purposes



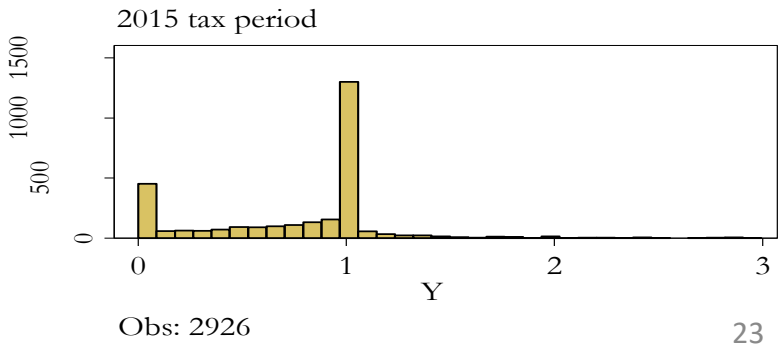
Obs: 3093



Obs: 2907



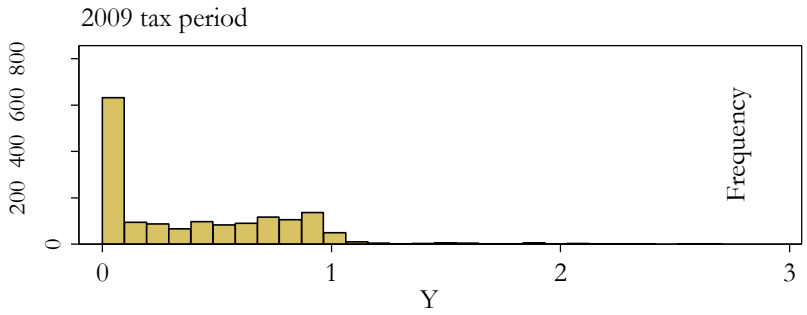
Obs: 2924



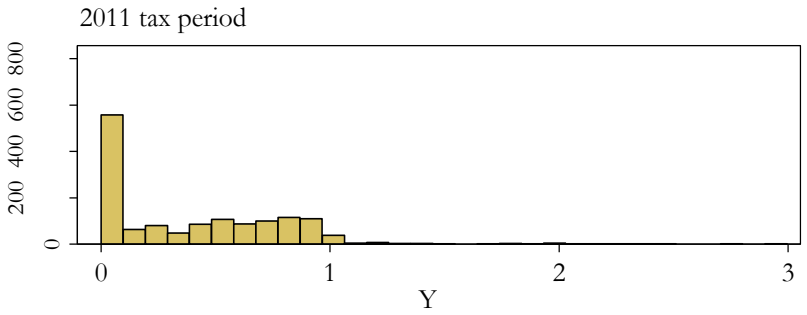
Obs: 2926

Descriptive statistics (2)

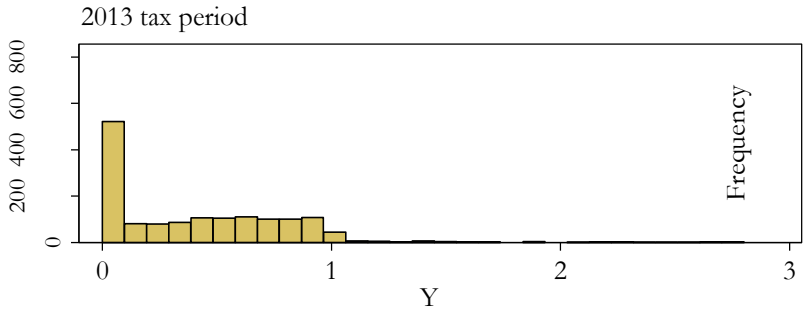
- Focusing on craftsmen, traders and farmers with INPS earnings equal to statutory thresholds, we see a marked reduction in the ratio between declared earnings for tax purposes and INPS earnings



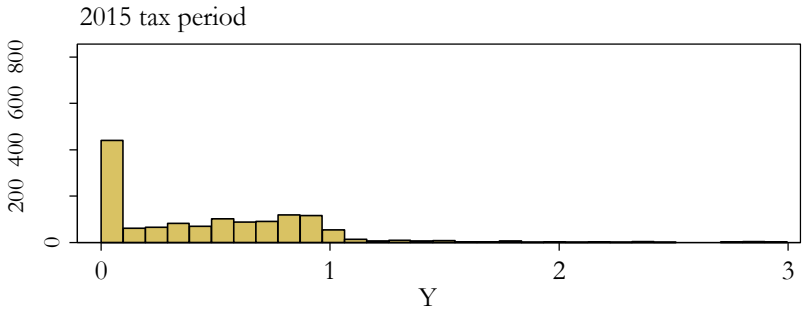
Obs: 1609



Obs: 1435



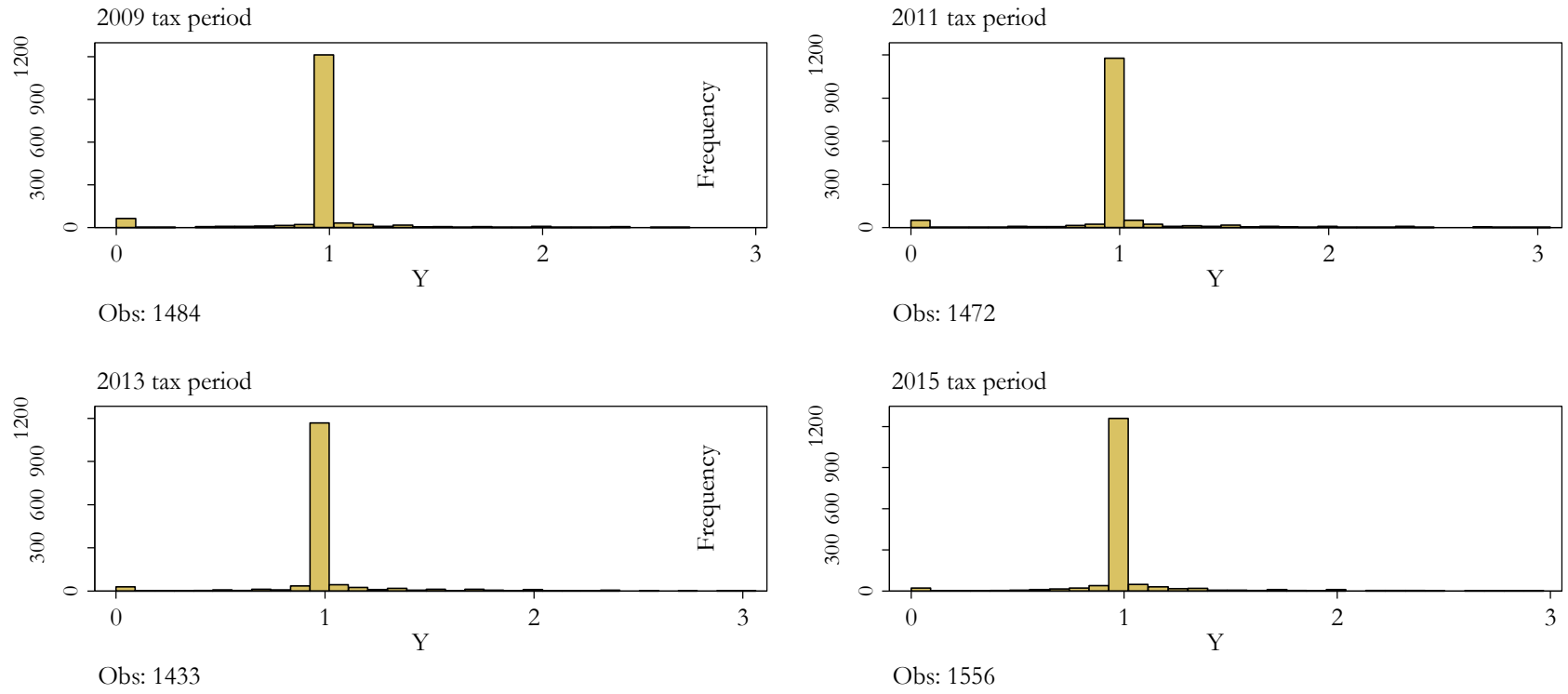
Obs: 1491



Obs: 1370

Descriptive statistics (3)

- Focusing on freelancers, we found high consistency between INPS earnings and tax returns data



Aim and method

- Aim: to adjust self-employment income for those with declared earnings for tax purposes lower than SICs-related income thresholds. This will avoid underestimation of poverty conditions among self-employed workers

- How? We dispose of 4 AD-SILC cross-sectional waves matched with tax returns data out of 14 (2009, 2011, 2013 and 2015 tax periods), where we can calculate the ratio between earnings as declared for tax purposes and earnings as collected in INPS archives

- Method: imputation with propensity score matching (Mahalanobis distance measure) dividing both 'donors' and 'treated' by 32 cells

Donors and treated (1)

- Who are the donors?

  We have selected among self-employed workers:

  i. Those for whom we know the ratio between earnings as declared for tax purposes and earnings as collected in INPS archives
  ii. Those craftsmen, traders and farmers who are exclusively self-employed workers
  iii. Those with INPS earnings equal to one of the 19 SICs-related thresholds that we have found in INPS archives (12 for craftsmen and traders; 7 for farmers) and farmers enrolled in the *Gestione Separata*
  iv. Those with age in the interval 16-79 and not in education

- Who are the treated?

  Self-employed workers included in the AD-SILC panel who meet the same requirements *ii, iii* and *iv* above, but not *i*.

- As a result:
  - The DONOR dataset is made of 5,821 observations

    71.2% of individuals repeated once throughout the period 2009-2015; 28.7% repeated twice; 0.1% repeated three times
  - The TREATED dataset is made of 23,381 observations

    46.7% individuals repeated once throughout the period 2004-2017; 29.5% repeated twice; 16.1% repeated three times; 7.7% repeated 4 times

Donors and treated (2)

- Cells:
  - Craftsmen and traders by SICs-related threshold (from 1 to 11 months worked) and sex: 22 cells
  - Craftsmen and traders by SICs-related threshold (12 months worked), sex and macro area (North West; North East; Middle; South): 8 cells
  - Farmers (regardless of the days of contribution accrued) by sex: 2 cells

craftsmen and traders

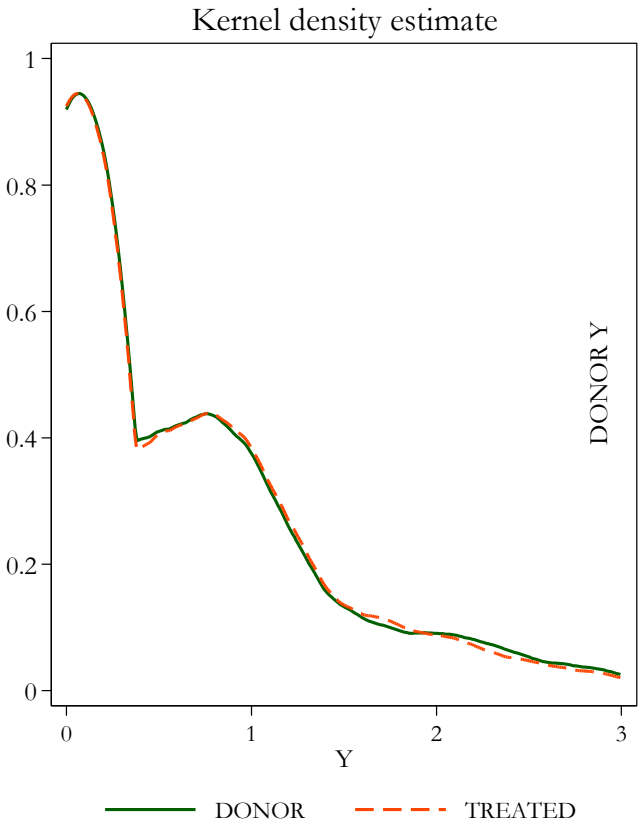
	DONOR		TREATED	
Months	Freq	*%*	Freq	*%*
12m	3846	*83.6*	14132	*79.8*
11m	70	*1.5*	312	*1.8*
10m	70	*1.5*	300	*1.7*
9m	118	*2.6*	530	*3.0*
8m	56	*1.2*	241	*1.4*
7m	48	*1.0*	328	*1.9*
6m	87	*1.9*	456	*2.6*
5m	44	*1.0*	211	*1.2*
4m	51	*1.1*	229	*1.3*
3m	111	*2.4*	491	*2.8*
2m	49	*1.1*	216	*1.2*
1m	53	*1.2*	259	*1.5*
Total	4603	*100.0*	17705	*100.0*

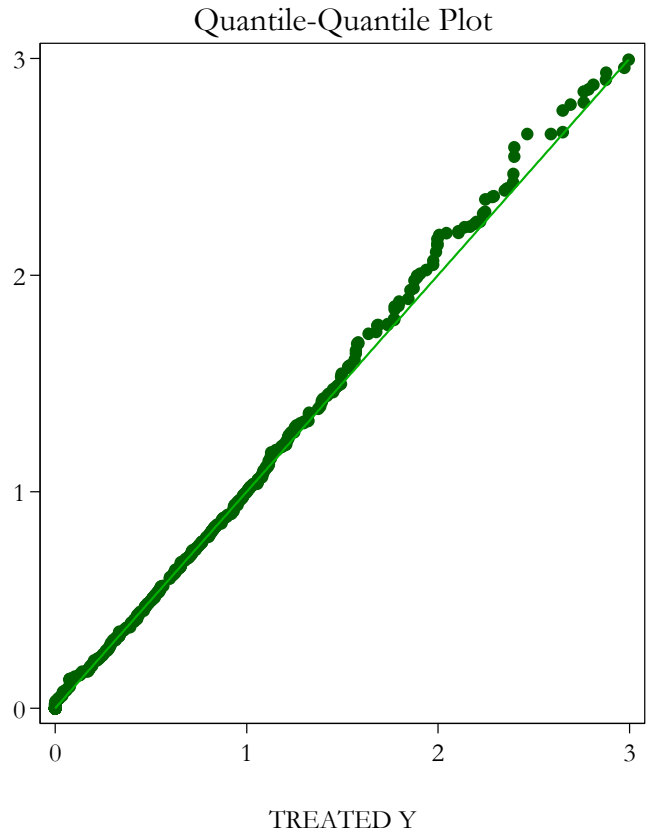
farmers

	DONOR		TREATED	
Days	Freq	*%*	Freq	*%*
156gg (iv)	67	*5.5*	265	*4.7*
156gg (iii)	105	*8.6*	400	*7.0*
156gg (ii)	21	*1.7*	85	*1.5*
156gg (i)	335	*27.5*	1765	*31.1*
208gg	446	*36.6*	1984	*35.0*
260gg	102	*8.4*	541	*9.5*
312gg	114	*9.4*	389	*6.9*
Other	28	*2.3*	247	*4.4*
Total	1218	*100.0*	5676	*100.0*

Results: marginal distributions (1)

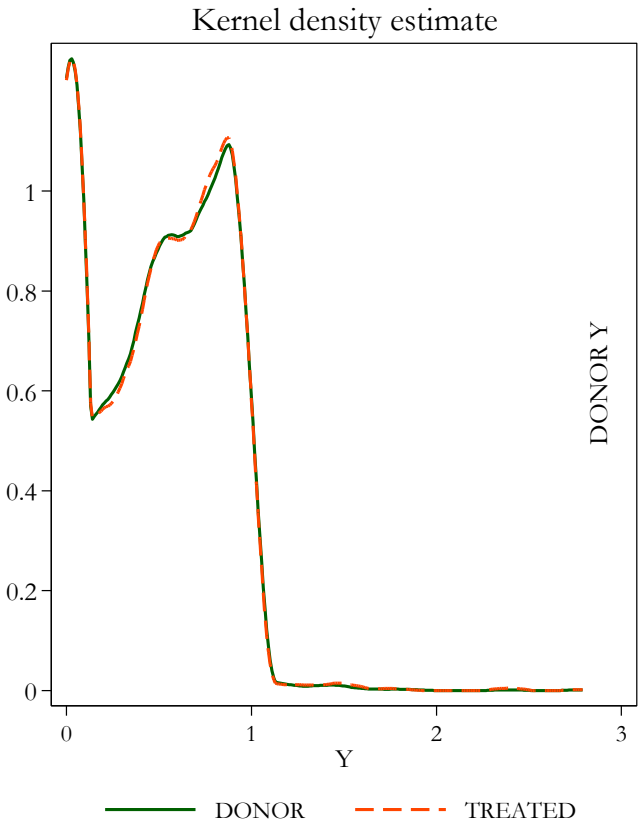
- Marginal distributions for craftsmen and traders (from 1 to 11 months worked)

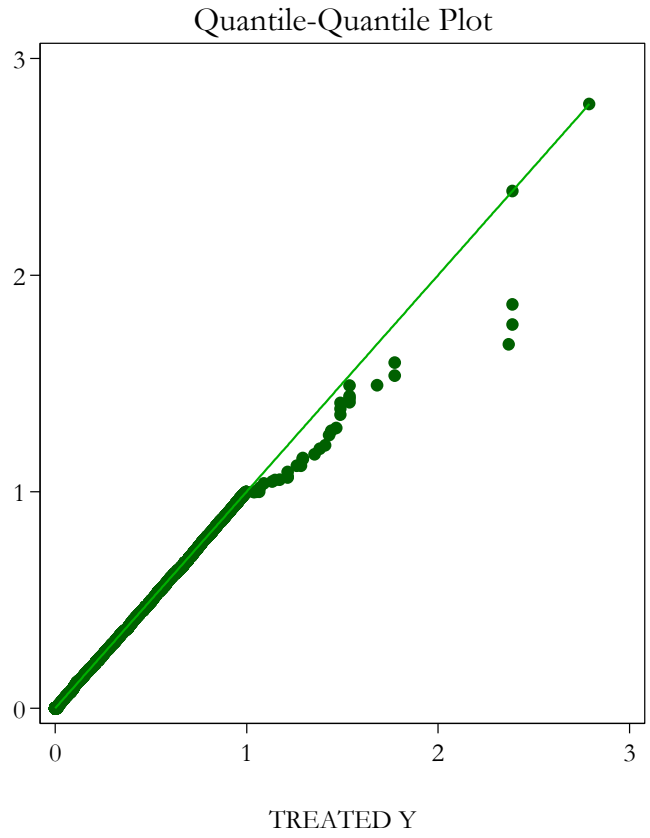




Results: marginal distributions (2)

- Marginal distributions for craftsmen and traders (12 months worked)





After the imputation

- We have imputed the ratio between declared earnings for tax purposes and INPS earnings for the AD-SILC observations included in the TREATED dataset

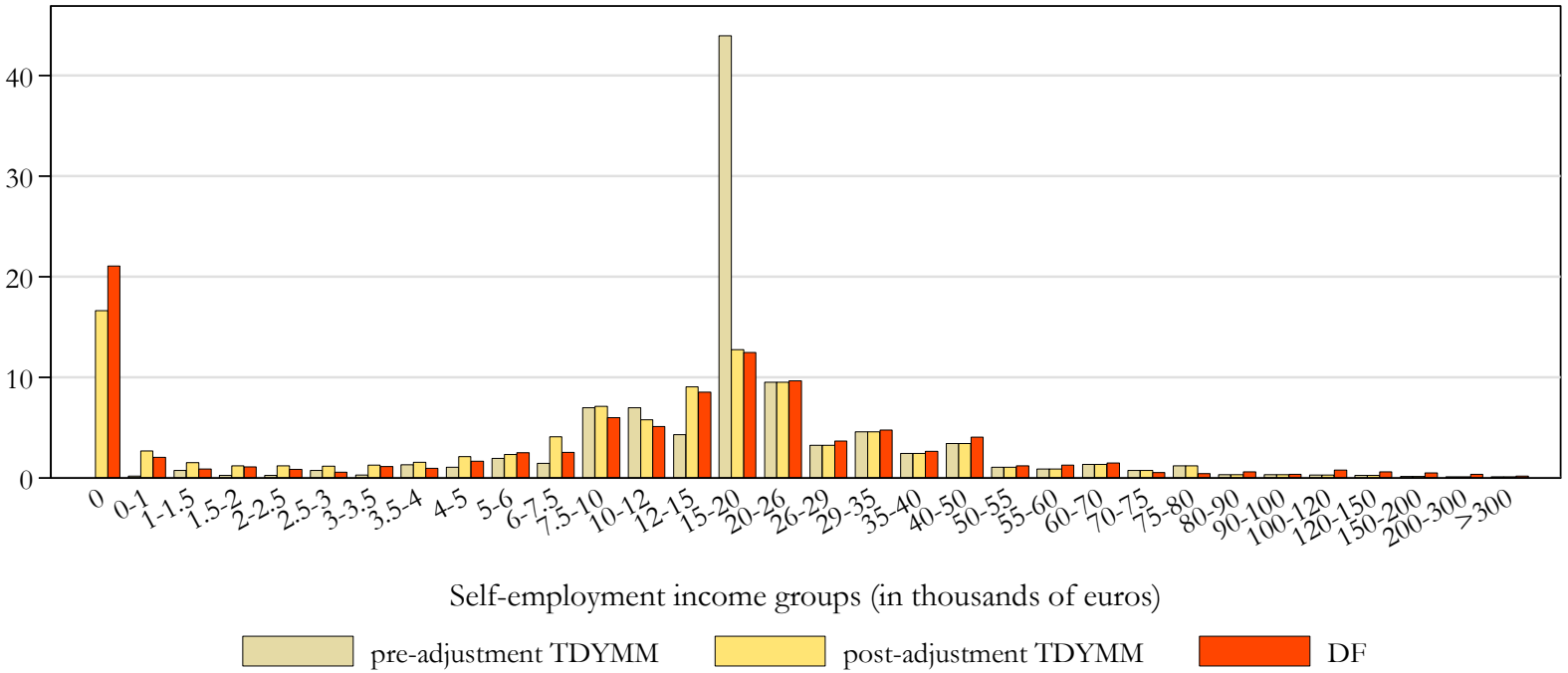
- Keep in mind the aim of the analysis (i.e. to avoid underestimation of poverty conditions):
  - if Y<=1, then we use imputed ratios to obtain estimated declared earnings
  - If Y>1, then we do not adjust earnings as collected in INPS archives

- We are unable to say whether declared earnings in the DONOR dataset are affected by tax evasion (quite likely), but we want INPS earnings to be a good approximation of administrative earning records for tax purposes

- Potential issues:
  - preservation of joint distributions (further validity tests are needed)
  - persistency of the phenomenon (Y_{t-2} might be a good predictor of Y)

Before and after the adjustment

- Number of self-employed workers before and after the adjustment by income groups



Note: non-weighted percentages
DF stands for tax returns micro data for the 2015 tax period
pre- and post-adjustment TDYMM were obtained from the AD-SILC dataset (2004-2017)

Weight calibration (1)

- Background:
  - Survey reweighting is a common practice in microsimulation studies
  - The aim is to improve the overall representativeness of a series of dimensions we are interested in with regard to T-DYMM's base year (2015, i.e. the year 0 of our simulations)
  - The base year's sample is the 2016 IT-SILC cross-sectional wave matched with INPS archives, tax returns data, cadastral data, and SHIW
  - We applied Deville and Särndal (1992)'s generalised raking procedure using the *sreweight* Stata command

Weight calibration (2)

- What are the dimensions included in the reweighting algorithm?
  - Dimensions used in the calibration of 2016 IT-SILC weights:
    - Distribution of the population by sex and fourteen 5-years age-groups at NUTS I level
    - Distribution of the population by sex and five age-groups at NUTS II level
    - Distribution of non-national population at NUTS I level by sex
    - Distribution of the population by demographic size of the municipality at Nuts I level
    - Number of households at NUTS II level

Weight calibration (3)

- What are the dimensions included in the reweighting algorithm?
  - Other dimensions we are interested in (such as):
    - Distribution of non-national population by sex, area of birth (EU and non-EU), and educational level
    - Distribution of households by number of family members
    - Distribution of the recipients of retirement income by sex and type of pensions (e.g. old-age pensions, survivors' pensions, disability pensions, and so on)
    - Distribution of recipients of specific income sources (e.g. rental income subject to *cedolare secca*; self-employment income subject to substitute tax regimes; and others)
    - Distribution of the population by age group and civil status
    - Distribution of the recipients of gross income subject to PIT by income groups
    - Distribution of the recipients of employment (retirement) income subject to PIT by income groups

Expansion and sampling

- Alignments and sample representativeness issues: subsequent to the weight calibration, we expanded the starting sample by multiplying individuals by calibrated weights; we drawn with replacement 100 sample of 200,000 households and selected the best-fitting sample

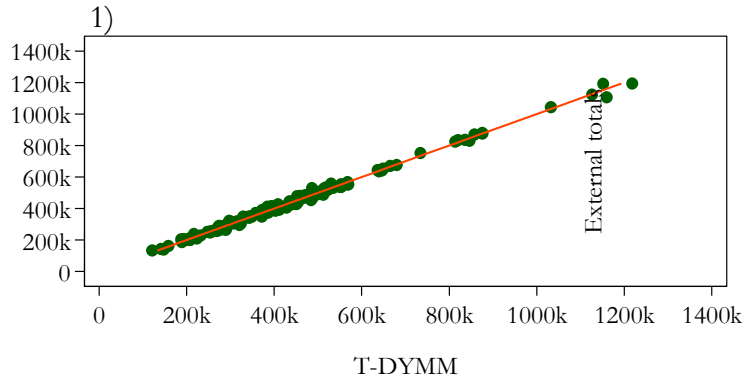
- As a result of this procedure, the starting sample is made of 477,643 individuals and 200,000 households

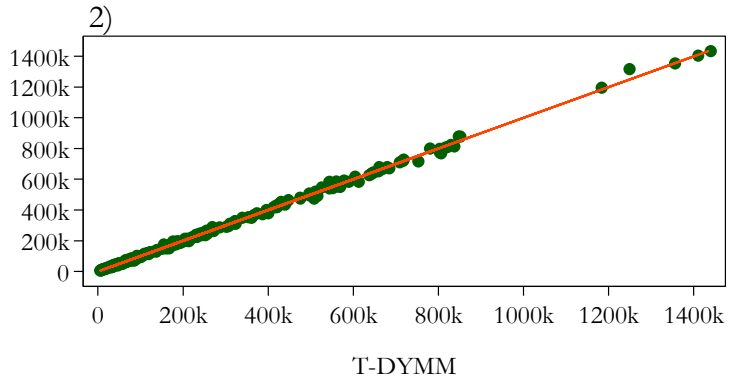
- Why 200,000 households? Sample dimension that ensures representativeness of cloned migrants

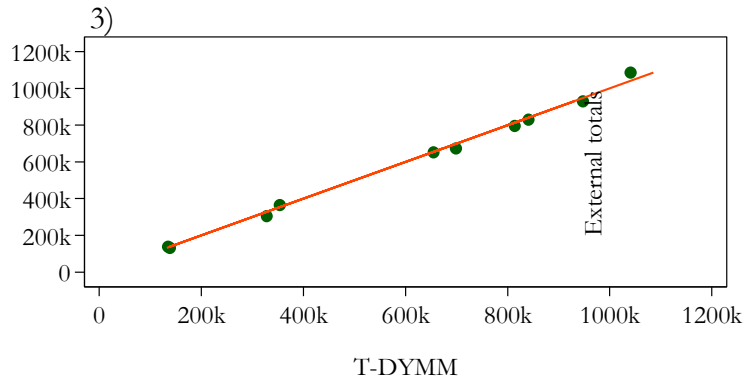
Results (1)

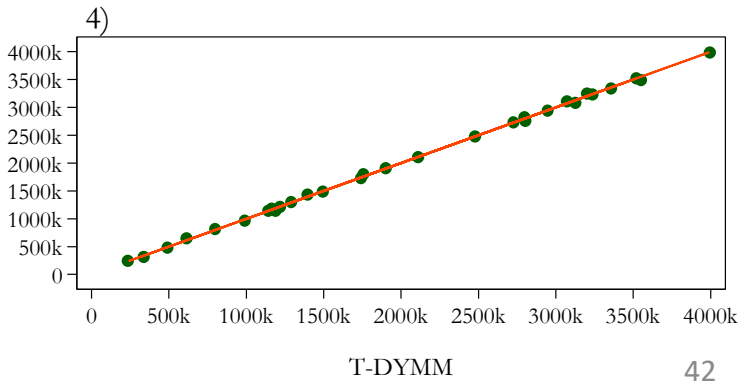
- Representativeness is preserved at the individual level

1) Distribution of the population by sex and fourteen 5-years age-groups at NUTS I level
2) Distribution of the population by sex and five age-groups at NUTS II level
3) Distribution of non-national population at NUTS I level by sex
4) Distribution of the population by demographic size of the municipality at Nuts I level





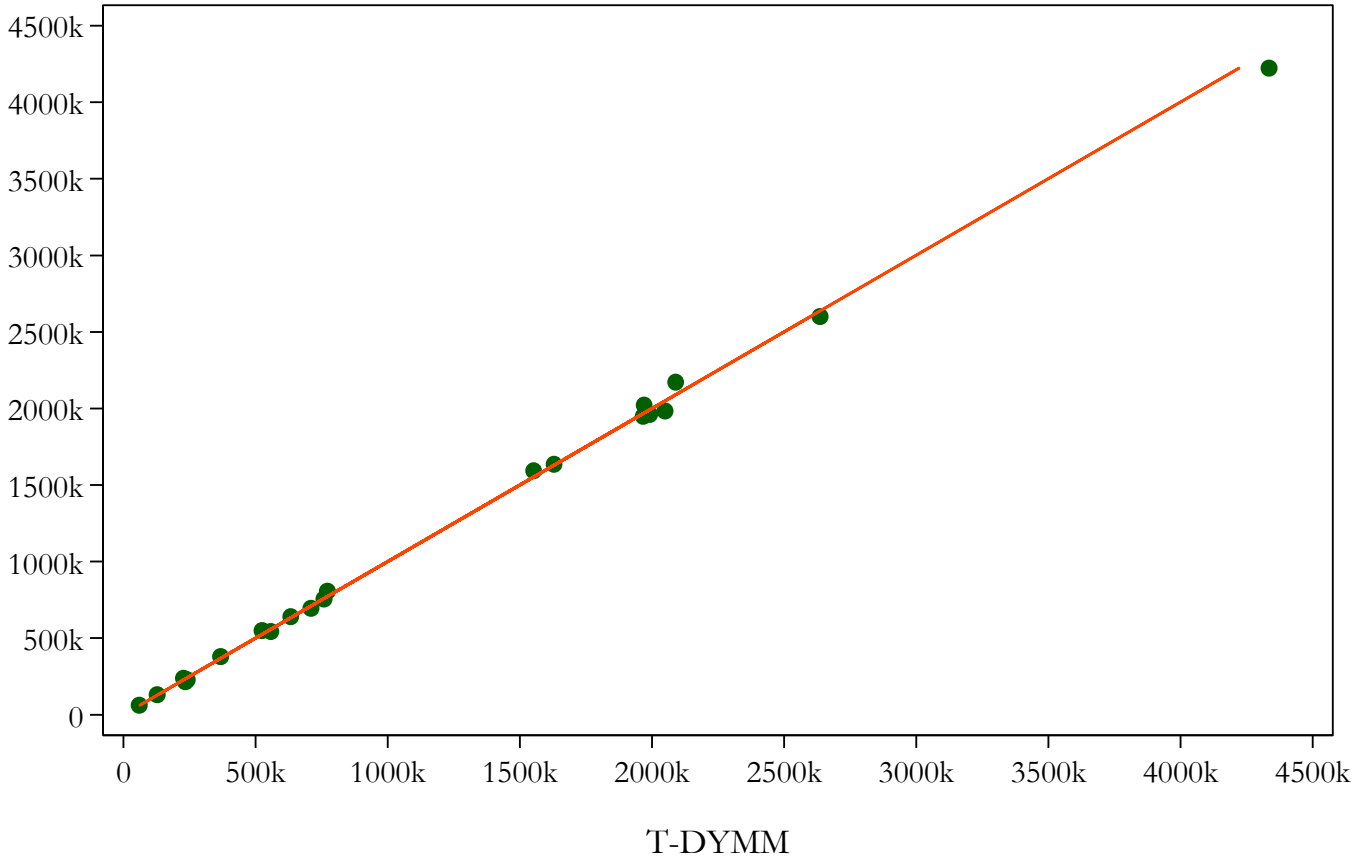




Results (2)

- Representativeness is preserved at the household level

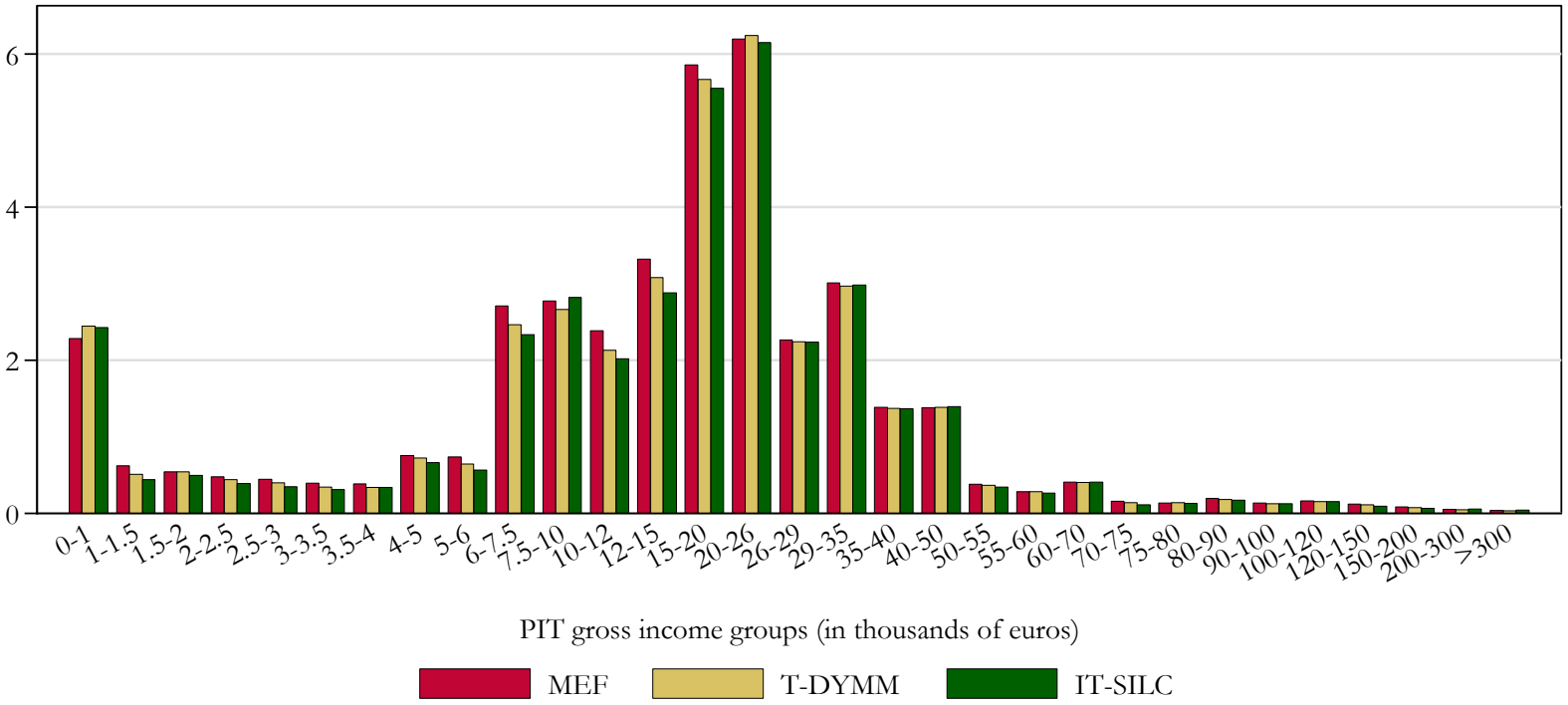
Number of households at NUTS II level



Results (3)

- Marginal improvements in each category of taxpayers and total income closer to the true total

Distribution of taxpayers with PIT gross income>0 by gross income groups

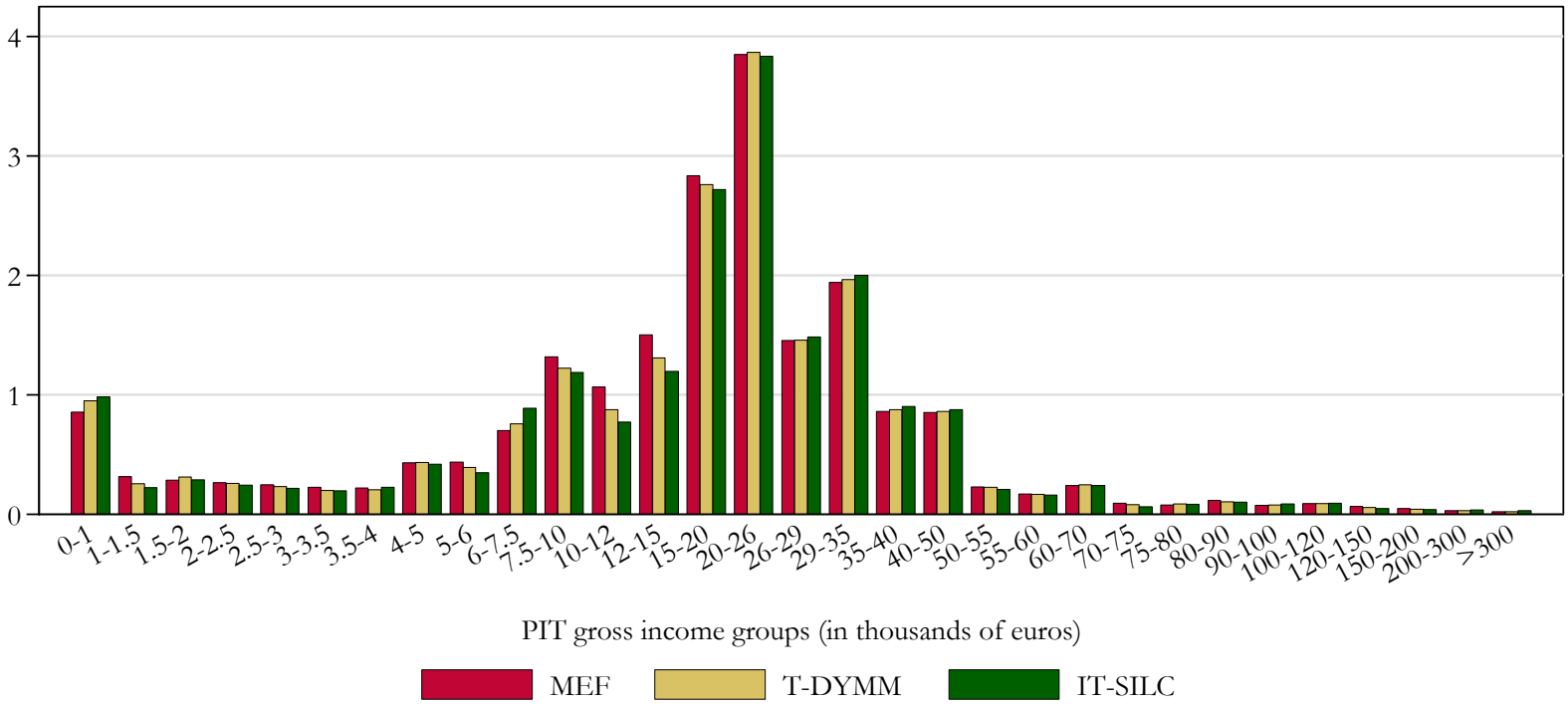


Taxpayers: MEF (40.0 millions); T-DYMM (38.6 millions); IT-SILC weight (37.6 millions)
PIT gross income: MEF (821.9 billions); T-DYMM (811.5 billions); IT-SILC's weight (801.7 billions)

Results (4)

- Marginal improvements in each category of taxpayers and total income closer to the true total

Distribution of taxpayers with employment income>0 by gross income groups

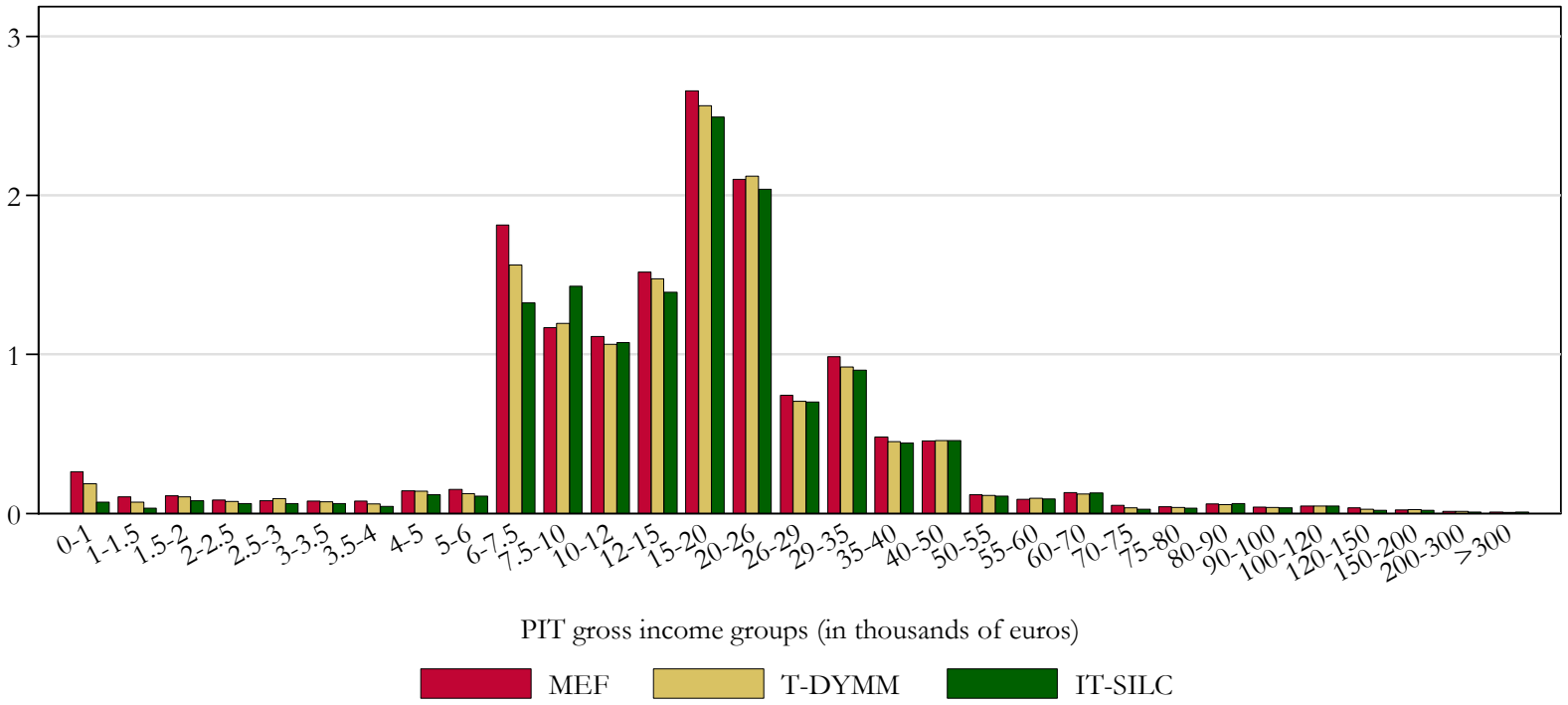


Taxpayers: MEF (20.9 millions); T-DYMM (20.4 millions); IT-SILC weight (20.2 millions)
Employment income: MEF (431.4 billions); T-DYMM (436.9 billions); IT-SILC weight (445.0 billions)

Results (5)

- Marginal improvements in each category of taxpayers and total income closer to the true total

Distribution of taxpayers with retirement income>0 by gross income groups

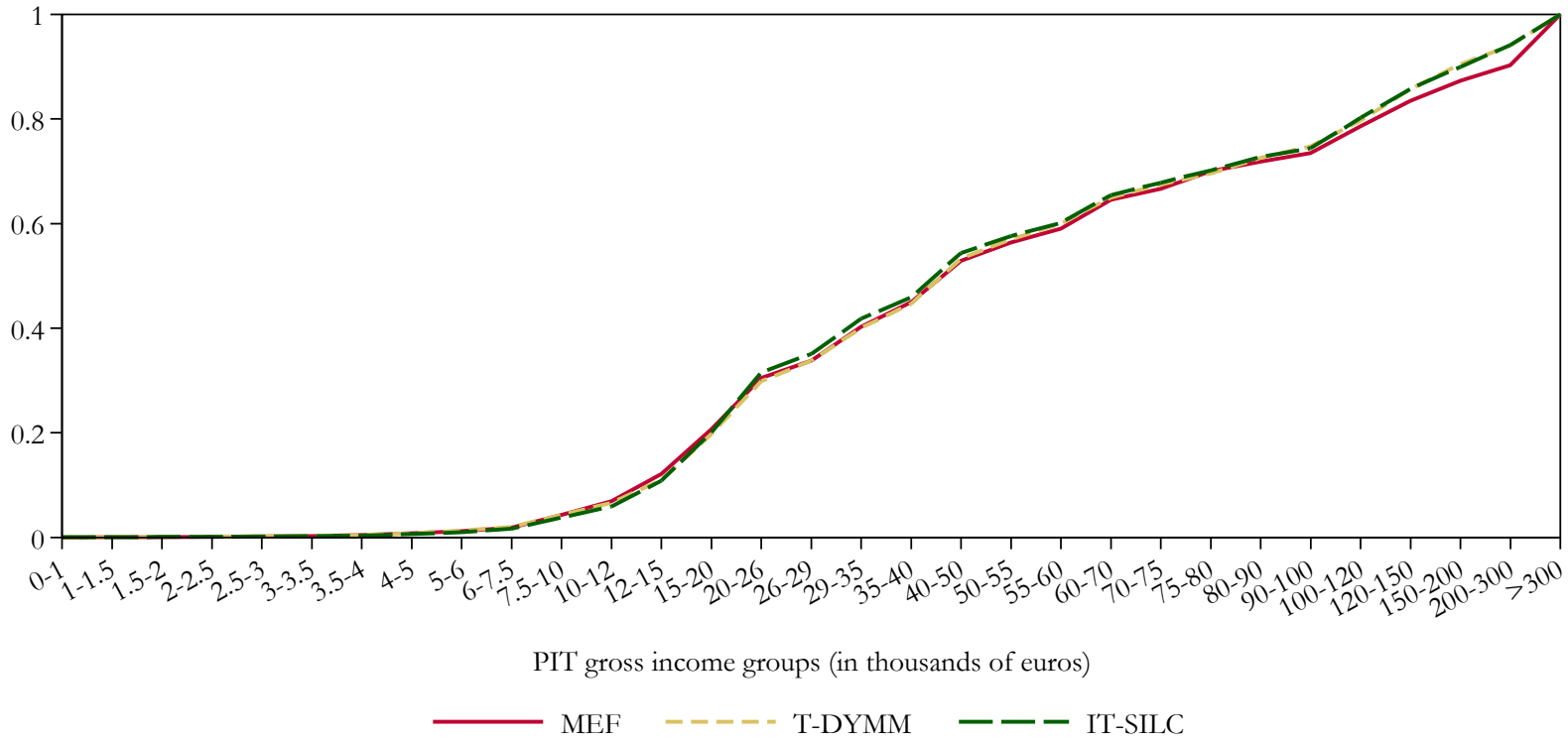


Taxpayers: MEF (14.8 millions); T-DYMM (14.0 millions); IT-SILC weight (13.5 millions)
Retirement income: MEF (249.2 billions); T-DYMM (245.5 billions); IT-SILC weight (239.6 billions)

Results (6)

- Total income closer to the true total

Cumulative self-employment income subject to PIT by gross income groups

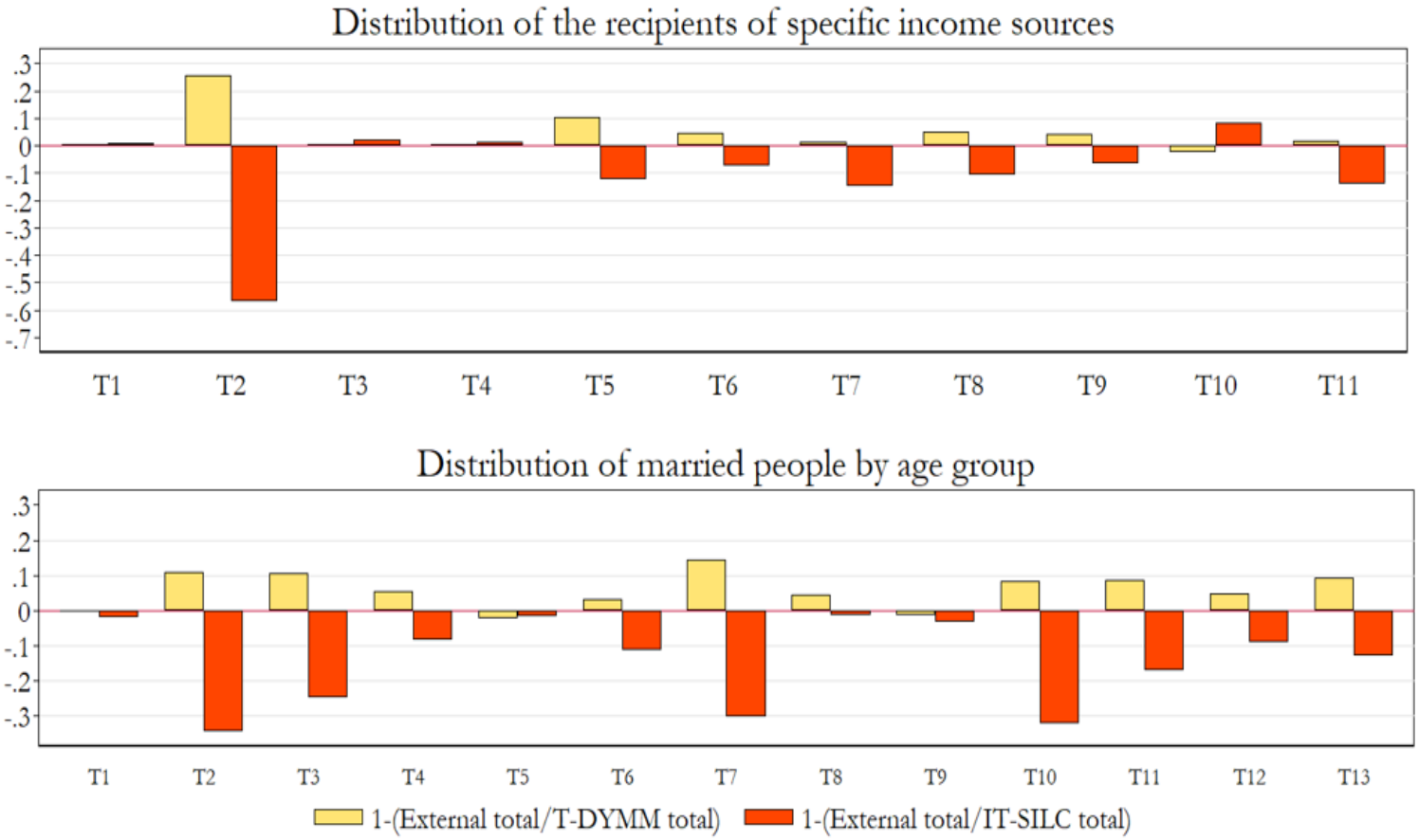


Self-employment income: MEF (102.0 billions); T-DYMM (94.3 billions); IT-SILC weight (80.4 billions)

Results (7)


Distribution of non-national population by sex, area of birth and eduacational level
.3
.2
.1
0
-.1
-.2
-.3
-.4
T1 T2 T3 T4 T5 T6 T7 T8 T9 T10 T11 T12
Distribution of households by number of family members
.2
.1
0
-.1
-.2
-.3
T1 T2 T3 T4 T5 T6
Distribution of the recipients of retirement income by type of pension
.2
.1
0
-.1
-.2
-.3
T1 T2 T3 T4 T5 T6 T7 T8 T9 T10 T11 T12 T13 T14
1-(External total/T-DYMM total)
1-(External total/IT-SILC total)

Results (8)


Distribution of the recipients of specific income sources
.3
.2
.1
0
-.1
-.2
-.3
-.4
-.5
-.6
-.7
T1
T2
T3
T4
T5
T6
T7
T8
T9
T10
T11
Distribution of married people by age group
.3
.2
.1
0
-.1
-.2
-.3
T1
T2
T3
T4
T5
T6
T7
T8
T9
T10
T11
T12
T13
1-(External total/T-DYMM total)
1-(External total/IT-SILC total)

References

Denis Chénard (2000). Individual alignment and group processing: an application to migration processes in DYNACAN, in Mitton L, H Sutherland and M Weeks (Eds.) *Microsimulation modelling for policy analysis: challenges and innovations*, Cambridge: Cambridge University Press, 238-247.

Gijs Dekkers (2015). On the modelling of immigration and emigration using LIAM2. 10.13140/RG.2.1.4373.8967.

Jean-Claude Deville and Carl-Erik Särndal (1992). Calibration estimators in survey sampling. *Journal of the American Statistical Association*, 87: 376-382.

MOSPI
Social protection systems