World Inequality Lab – Technical Note N° 2021/11

DINA income series for Italy, 2004-2015

Demetrio Guzzardi
Elisa Palagi

November 2021

WORLD INEQUALITY LAB

WID.WORLD
THE SOURCE FOR GLOBAL INEQUALITY DATA

DINA income series for Italy, 2004-2015: A technical note

Demetrio Guzzardi* Elisa Palagi†

1 Methodology

In this section we describe the methodology used in our companion paper "Income inequality, effective tax rates and the progressivity of the tax system: Evidence from Italian Distributional National Accounts" (Guzzardi, Palagi, Roventini and Santoro (2021)) to estimate new series of income inequality for Italy following the DINA guidelines.

To achieve this result, we combine several data sources such as National surveys, National Accounts, Regional Accounts, Personal Income Tax returns, and external data on Wealth distribution[1]. We start by using the IT-SILC survey as our database of reference due to an acceptable level of detail on many income sources reporting both net and gross variables. To correct for non-sampling errors from which the IT-SILC is suffering, we recalibrate the survey sample weights using the Personal income tax tabulations at the regional level. Thereafter, using data-fusion techniques, we use the SHIW survey from the Bank of Italy to derive the joint distribution of (i) wealth and income and (ii) consumption and income at the personal level. With this information on the two joint distributions we can therefore integrate our

*Institute of Economics and EMbeDS, Scuola Superiore Sant'Anna, Pisa, Italy; E-mail address: demetrio.guzzardi@santannapisa.it

†Institute of Economics and EMbeDS, Scuola Superiore Sant'Anna, Pisa, Italy; E-mail address: elisa.palagi@santannapisa.it

[1]The use of all the data sources involved takes place under the full and sole responsibility of the authors and does not involve the institutions providing the data.

dataset with the best available data on consumption and wealth, respectively the HBS survey on consumption, and a novel data source on Italian Wealth distribution created by Acciari, Alvaredo, and Morelli (2020). Finally, using National and Regional Accounts to rescale income sources and taxes to match their macroeconomic counterparts, we construct National Income distribution series. Let us analyze each step in detail.

1.1 Rescaling the sampling weights

Several studies (Dalenius 1977; Assael and Keon 1982; Gertner and Köhl 1992; Verma and Lê 1996; Taleb and Douady 2015) show that National surveys typically suffer from sampling and non-sampling errors. Sampling errors are statistical errors that could potentially be solved with a large enough sample size. In particular, due to a small sample size, surveys may underestimate the total income owned by a specific group of individuals. This is especially true at the top of the income distribution, where revenues are often under-reported or misreported. To overcome these issues in the IT-SILC, the National Statistical Institute identifies survey respondents by their fiscal code to match their income with external administrative data. In this way, misreporting of several income items can be corrected with remarkable precision for wages, pensions, and other transfers.

Non-sampling errors, on the other hand, are errors that cannot be solved by increasing the sample size, and that typically arise due to unobserved heterogeneity in non-response rates. The construction of the IT-SILC sample-weights considers the non-response rates of individuals by matching for each non-respondent the equivalent respondent based on several demographic characteristics and occupation. However, it has been noted that the non-response rates may increase with higher income (Groves and Couper 2012). Therefore, not considering the totality of income in the construction of the sample weights leads to biased results by under-representing the richest individuals and over-representing those at the bottom of the income distribution. Recently, the national statistical office has acknowledged this issue (ISTAT 2021) and has considered possible ways to account for these types of non-sampling

errors using administrative data to fill the reported income of non-respondents. However, the publicly available data have not been adjusted yet. To overcome these problems, we therefore employ a new algorithm developed by Blanchet, Flores, Morgan, et al. (2018), which uses tax tabulations to correct the sample weights. The algorithm identifies the merging point between the income distribution derived from survey and tax data and rescales the sample weights at the right of the merging point to match the frequencies of the tax data, which is assumed to be more reliable. To compensate for the scaling-up at the top, the algorithm also scales down the weights at the left of the merging point to keep total sample weights constant. Moreover, this algorithm allows preserving the original distribution of several covariates such as age and gender.

This technique has been recently used in Blanchet, Chancel, and Gethin (2019), also for the case of Italy. However, we made several adjustments: (i) we used a definition of taxable income that is more in line with the income reported in income tax returns; (ii) we only aim at correcting non-sampling errors since the IT-SILC already corrects for the possible sampling errors, as explained above; (iii) we use regional tax tabulations to correct each region's non-sampling error and keep the original distribution of gender and age at the regional level (full detail in appendix A.1).

1.2 Data fusion

Although rich in information on income and demographic covariates, the IT-SILC completely misses information on wealth and consumption behaviour[2]. However, these elements are fundamental to distribute income and taxes linked to financial assets, real estates, and consumption. To acquire reliable information on the distribution of consumption, we use the Household and Budget Survey (HBS) produced by IStat, while we use a novel data source by Acciari, Alvaredo, and Morelli (2020) (henceforth AAM), estimated upon National Accounts and administrative data on inheritance taxation, to gather information on wealth

[2]Income from financial assets is present in the IT-SILC, but it is severely under-reported. It represents less than 10% of the capital income from NA.

distribution.

To combine these datasets, we first obtain the joint distribution of wealth and income and of consumption and income from The Survey on Household Income and Wealth (SHIW) by the Bank of Italy, which contains information on income, wealth and consumption. In SHIW, both wealth assets and consumption are recorded at the household level. Therefore, we first redistribute wealth at the personal level following the methodology developed in D'Alessio (2018) and allocate consumption among the family members simply in proportions of their net income. Similarly to Albarea et al. (2015), we merge the two surveys SHIW and IT-SILC by propensity score matching using wages, self-employment income, pensions, gender, age and geographical area as covariates for the matching algorithm (full detail in appendix A.2).

At this point, we have an IT-SILC survey with additional data on wealth ownership and consumption from SHIW, which identifies the joint distributions between income and wealth and between income and consumption, which is a crucial piece of information to investigate the overall progressivity of a tax system (Kuypers, Figari, and Verbist 2018).

To integrate these external data sources on the distribution of wealth and consumption into our main data, we proceed as follows. First, we rank each person by percentiles of wealth; then, we associate to each percentile the wealth share corresponding to the same percentile of wealth derived from AAM data. Finally, using the total national wealth calculated by AAM and multiplying it by the shares of each percentile, we can derive the whole distribution of wealth consistent with AAM. Moreover, we further decompose the total net wealth in six different components using the composition of wealth in SHIW at the percentile level (full detail in appendix A.2).

For the case of consumption, we apply an analogous procedure. We first sum at the family level the personal consumption and rank it by one thousand fractiles of consumption. We then use the HBS to derive the distribution of consumption at the same fractile level, and we apply to each consumption-fractile in the IT-SILC the same level of consumption derived from the HBS.

1.3 Deriving the Distributional National Accounts

We have derived an IT-SILC survey with recalibrated sample weights that is augmented with data on wealth and consumption. Following the DINA guidelines (Alvaredo et al. 2016), we use this database to estimate the distribution of each income component and taxes that constitute the national income in National Accounts. We thus identify four income concepts: Factor National Income, Pre-tax National Income, Post-tax disposable Income and Post-tax National income.

In order to construct series for the different income definitions, we first identify in the National Accounts all the direct taxes paid by households and distribute them in proportion to the original distribution of the relative direct taxes paid by individuals identified in IT-SILC[3]. We then add to the net variable this final amount of direct taxes paid to obtain a provisional amount of gross income for each income category. From the Regional National Accounts, we identify the income from self-employment, wages, actual and imputed rents, dividends and interests for the institutional sector of households. We distribute them in proportion to their relative provisional gross income keeping the regional totals consistent with Regional National Accounts, thus obtaining what we call the final gross income. To allot actual and imputed rents, as well as dividends and interests, we use the distribution of real estates, financial equities and shares. With this approach, we implicitly assume that the rate of return on each asset is constant over the wealth distribution. This is a relatively strong assumption, as recent findings for other countries points out that higher levels of wealth are associated to higher rates of returns (Fagereng et al. 2020; Bach, Calvet, and Sodini 2020; Iacono and Palagi 2021). However, we keep this assumption, as it is standard in similar studies in the literature (see Piketty, Saez, and Zucman (2018) for the US case). Therefore, it is worth noticing that we are probably underestimating the financial and estate income accruing to the wealthiest individuals and thus reducing the overall level of inequality.[4] We

[3]The direct taxes are calculated in the survey as the difference between gross and net variables

[4]We choose to keep the assumption of constant returns in order to obtain results that are comparable to previous studies for other countries (Piketty, Saez, and Zucman 2018) and due to the lack of estimates on

estimate the net income variables as the difference between the final gross income variables and the final direct taxes paid. Finally, we distribute indirect taxes on consumption in proportion to personal consumption and distribute other indirect taxes on income or wealth proportionally to the relative income or wealth asset (full detail in Appendix A.3).

To match the National Income of the whole economy, we need to include also the income accruing to the Public and Business sectors of the National Accounts. Following the literature, we consider the corporate tax and retained earnings of the business sector as income earned by those who own the shares of the businesses. Therefore, we distribute the latter categories in proportion to the financial assets of equities and shares. Regarding the income from the public sector, we distribute it in proportion to the personal distribution of all other income sources. This distributional choice operates as a level-shifter of individual income but will not change the relative distribution among individuals. In addition, to construct the Post-tax National income series, we distribute the public spending according to the actual expenditure across regions.

We then obtain a final dataset that is consistent with National and Regional Accounts and that distributes at the personal level all gross income variables, Social Security contributions, direct and indirect taxes.

heterogeneous rates of return specific for Italy. As such, our results should be read as conservative, and the true underlying inequality levels might be even more dramatic.

References

Acciari, Paolo, Facundo Alvaredo, and Salvatore Morelli (2020). "The concentration of personal wealth in Italy 1995-2016". In: *Measuring and Understanding the Distribution and Intra/Inter-Generational Mobility of Income and Wealth*. University of Chicago Press.

Albarea, Andrea, Michele Bernasconi, Cinzia Di Novi, Anna Marenzi, Dino Rizzi, and Francesca Zantomio (2015). "Accounting for tax evasion profiles and tax expenditures in microsimulation modelling. The BETAMOD model for personal income taxes in Italy". In: *University Ca'Foscari of Venice, Dept. of Economics Research Paper Series No* 24.

Alvaredo, Facundo, Anthony Atkinson, Lucas Chancel, Thomas Piketty, Emmanuel Saez, and Gabriel Zucman (2016). "Distributional National Accounts (DINA) guidelines: Concepts and methods used in WID. world". In:

Alvaredo, Facundo and Elena Pisano (2010). "Top incomes in Italy, 1974–2004". In: *Top Incomes: A Global Perspective*, pp. 625–63.

Assael, Henry and John Keon (1982). "Nonsampling vs. sampling errors in survey research". In: *Journal of Marketing* 46.2, pp. 114–123.

Bach, Laurent, Laurent E Calvet, and Paolo Sodini (2020). "Rich pickings? Risk, return, and skill in household wealth". In: *American Economic Review* 110.9, pp. 2703–47.

Blanchet, Thomas, Lucas Chancel, and Amory Gethin (2019). "How Unequal Is Europe? Evidence from Distributional National Accounts, 1980-2017". In: *WID. world working paper* 6.

Blanchet, Thomas, Ignacio Flores, and Marc Morgan (2018). "BFMCORR: Stata module for correcting surveys using tax data". In:

Blanchet, Thomas, Ignacio Flores, Marc Morgan, et al. (2018). "The weight of the rich: improving surveys using tax data". In: *WID. world Working Paper* 12.

Blanchet, Thomas, Juliette Fournier, and Thomas Piketty (2017). "Generalized Pareto curves: theory and applications". In: *Review of Income and Wealth*.

Dalenius, Tore (1977). "Bibliography on Non-Sampling Errors in Surveys: I (A to G)". In: *International Statistical Review/Revue Internationale de Statistique*, pp. 71–89.

D'Alessio, Giovanni (2018). "Gender wealth gap in Italy". In: *Bank of Italy Occasional Paper* 433.

Di Caro, Paolo et al. (2018). *Redistribution in real-world PIT: Evidence from Italian tax records*. Tech. rep.

Fagereng, Andreas, Luigi Guiso, Davide Malacrino, and Luigi Pistaferri (2020). "Heterogeneity and persistence in returns to wealth". In: *Econometrica* 88.1, pp. 115–170.

Gertner, G and M Köhl (1992). "An assessment of some nonsampling errors in a national survey using an error budget". In: *Forest Science* 38.3, pp. 525–538.

Groves, Robert M and Mick P Couper (2012). *Nonresponse in household interview surveys*. John Wiley & Sons. Chap. 5.

Guzzardi, Demetrio, Elisa Palagi, Andrea Roventini, and Alessandro Santoro (2021). *Income inequality, effective tax rates and the progressivity of the tax system: Evidence from Italian Distributional National Accounts*. Working Paper.

Iacono, Roberto and Elisa Palagi (2021). "A Micro Perspective on r >g". In: *WID.world working paper* 2021/03.

ISTAT. "Istat, Indagine sulle condizioni di vita (UDB - IT SILC)". In: ().

— "Istat, Indagine sulle spese delle famiglie: Microdati ad uso pubblico". In: ().

— (2021). "L'indagine EU-SILC. Innovazioni nella metodologia di rilevazione e di stima". In: *Letture statistiche – Metodi*, pp. 137–150.

Italy, Bank of. "Banca d'Italia, Indagine sui bilanci delle famiglie italiane - Version 10.1 of the historical database". In: ().

Kuypers, Sarah, Francesco Figari, and Gerlinde Verbist (2018). "Redistribution in a joint income–wealth perspective: A cross-country comparison". In: *Socio-Economic Review*.

MEF. "IRPEF tax tabulation - MEF - Dipartimento delle Finanze". In: ().

Piketty, Thomas, Emmanuel Saez, and Gabriel Zucman (2018). "Distributional national accounts: methods and estimates for the United States". In: *The Quarterly Journal of Economics* 133.2, pp. 553–609.

Taleb, Nassim Nicholas and Raphael Douady (2015). "On the super-additivity and estimation biases of quantile contributions". In: *Physica A: Statistical Mechanics and its Applications* 429, pp. 252–260.

Verma, Vijay and Thanh Lê (1996). "An analysis of sampling errors for the demographic and health surveys". In: *International Statistical Review/Revue Internationale de Statistique*, pp. 265–294.

A Methodological details

A.1 Rescaling the sampling weights

To correct the IT-SILC survey from non-sampling error, we use the "BFMcorrection" algorithm created by Blanchet, Flores, and Morgan (2018). This method requires the identification of a taxable income definition that is as close as possible to the one present in tax returns. By comparing the data from survey and tax returns, the algorithm, proceeds to recalibrate the sample weights of the survey by identifying the merging point between the distribution of income in the survey and the distribution of income in tax returns elaborated through the use of the Generalized Pareto interpolation method (Blanchet, Fournier, and Piketty 2017).

A similar procedure was performed by Blanchet, Chancel, and Gethin (2019) (henceforth BCG) in their study about European countries. However, in that case, the authors apply a definition of taxable income for Italy that is not entirely homogeneous with tax returns data.

Specifically, the authors use the following definition of income: Gross employee cash or near cash income (PY010G) + Company car (PY021G) + Gross cash benefits or losses from self-employment (including royalties) (PY050G) + Pensions received from individual private plans (other than those covered under ESSPROS) (PY080G) + Unemployment benefits (PY090G) + Old-age benefits (PY100G) + Survivor benefits (PY110G) + Sickness benefits (PY120G) + Disability benefits (PY130G) + Education-related allowances (PY140G) + Income from rental of a property or land (HY040G) + Interests, dividends, profit from capital investments in unincorporated business (HY090G) + Income received by people aged under 16 (HY110G).

From this income, they subtract the social contributions paid by employees and self-employed workers, estimated using OECD macro-aggregates. However, there are some inaccuracies in this definition. Not all the income they use is actually part of the total income reported on tax returns. Regarding social transfers, only transfers due to unemployment and

old-age pensions are subject to IRPEF. In contrast, transfers due to illness, disability etc., do not contribute to the formation of the total income for IRPEF purposes and, therefore, are not reported in the income tax return. In this study, therefore, we use a different definition of taxable income closer to the definition of the Italian tax system. We start from the income determined by BCG, but we also subtract Disability benefits (PY130G) and Education-related allowances (PY140G) as they are not taxable income. Regarding income received by people aged under 16 (HY110G), this counts for a small part. Still, in general, the incomes of the underaged must be included in the parents' tax return unless the under-aged are subject to legal usufruct. In that case, the under-aged's income must be declared in the under-aged's name. This eventuality is particularly rare, so for simplicity, we include the income received by people under 16 as income of the household's head. The so-called Fringe Benefits contribute to the formation of the employee's total income and are subject to IRPEF taxation if they exceed 258 euros (Article 51 paragraph 3 of the TIUR). Therefore we include them if they exceed this threshold as opposed to BCG, which only includes the variable company car (PY021G).

As for the social contributions, in BCG, they are approximated through OECD aggregates and both the contributions of self-employed workers and employees are excluded. We instead use the values of the variables already present in the IT-SILC ("csa" and "csdi"), including, however, contributions from self-employed workers in the definition of total income. Although they are deducted for the calculation of the taxable income, this type of contributions are still reported in tax returns and have to be part of the total income used for the correction of the sample weights. Regarding dividends and other capital income, only part of this income must be reported in tax returns; therefore, we follow Alvaredo and Pisano (2010) by correcting IRPEF tax return revaluing the capital income in tax return by 2.5.

Another important update with respect to previous studies is that we apply this correction at the regional level. In fact, we use the regional tax returns from 2004 to 2015 and rescale the sample weights individually for each one of the 21 regions present in the IT-SILC.

Furthermore, unlike BCG, we do not correct the survey through replacement income above the merging point since the statistical office already considers this type of error by making an individual matching with administrative data.

Finally, we constrain the correction of the weights to keep the gender and age covariates constant for each region.

In Table 1 we show for each variable the average amount, maximum value, standard deviation and total in the survey. Minimum values exists only for PY050G Gross cash benefits or losses from self-employment which can be negative due to losses. In 2015 the minimum value of PY050G was €-40,000.

Table 1: Descriptive statistics on SILC variable in 2015

Variable code	*Variable name*	*Mean*	*Max*	*Standard Deviation*	*Total, Millions*
py010g	Gross employee cash or near cash income	7,901	857,552	15,777	480,699
csdi	SSC employee	708	9,502	1,328	43,073
csda	SSC employeer	2,676	94,529	5,047	162,799
py021g	Company car	14	8,646	217	856
py020g	Fringe benefits	108	17,822	534	6,595
py050g	Gross cash benefits or losses from self-employment	2,937	390,428	12,481	178,689
csa	SSC self-employed	489	27,754	2,088	29,771
py080g	Pensions received from individual private plans	2	9,882	123	132
py090g	Unemployment benefits	417	161,161	2,911	25,342
py100g	Old-age benefits	3,582	280,708	9,312	217,925
py110g	Survivor benefits	736	76,537	3,285	44,793
py130g	Disability benefits	226	90,133	1,849	13,733
py140g	Education-related allowances	26	54,000	712	1,572
hy040g	Income from rental of a property or land	431	181,285	3,154	26,238
hy090g	Interests, dividends, profit from capital investments	167	21,927	811	10,172
hy110g	Income received by people aged under 16	2	10,207	90	110

A.2 Data fusion

The IT-SILC data released by IStat are part of the European EU-SILC. They include the target variables of the EU-SILC survey, but also the detail of social contributions paid by employees and self-employed workers. It is an annual survey and is representative both at the

national and regional level. This allows us to have an additional level of detail to distribute national income. Nevertheless, although very rich in terms of income variables, the survey does not have information about the consumption of households or about financial and real estate properties, which are fundamental for estimating the distribution of taxes.

The Survey on Household Income and Wealth (SHIW) is carried out by the Bank of Italy every two years and is representative of the population residing in the macro areas of the country. Although SHIW has less detailed information on household incomes, it has a lot of information on financial, real estate assets and consumption. Therefore, we intend to build an augmented IT-SILC survey, with data on wealth added using the joint distribution between income and wealth and between income and consumption present in the SHIW survey. To merge the two surveys, we use a propensity score matching method with Mahalanobis distance between income, age, macro-region and gender, similarly to the method in Albarea et al. (2015), with the difference that in our case, we proceed with matching on a personal rather than a family level. To proceed with the propensity score matching, we have aggregated the incomes in SILC and SHIW following a comparable definition of income given by the sum of income from employment, self-employed, pensions, and other transfers. In SHIW we have combined: Compensation of Employee (yl) + Pensions and other transfers (yt) + Net income from self-employment and entrepreneurial income without profits and dividends (ym-ym3). In SILC, on the other hand, after having divided family incomes equally among the adult members of the family, we have combined: Employee cash or near cash income (py010n) + Fringe benefit (py020n) + Cash benefits or losses from self-employment (py050n) + Pension from individual private plans (py080n) + Unemployment benefits (py090n) + Old-age benefits (py100n) + Survivor benefits (py110n) + Sickness benefits (py120n) + Disability benefits (py130n) + Education-related Allowances (py140n) + Family / children-related allowances (hy050n split) + Social exclusion not elsewhere classified (hy060n split) + Housing allowances (hy070n split). At this point, we proceed with joining the two surveys using propensity score matching with Mahalanobis distance to obtain a single database with

SILC data but with the addition of data on real estate, property and consumption.

However, both consumption and wealth variables that are present in SHIW are recorded at the family level. Therefore, we distribute wealth among family members following D'Alessio (2018)'s methodology, while we allocate consumption to family members in proportion to personal income. In addition, since the SHIW is carried out every two years, we add the missing year in SHIW using the same data from the previous year. In this way, we obtain an annual SHIW questionnaire to be matched with the SILC questionnaire.

Acciari, Alvaredo, and Morelli (2020) (hereinafter AAM) observed that the distribution of wealth deriving from SHIW data is significantly less concentrated than reported in their study, which uses data from administrative sources to distribute the National Wealth at the individual level. Therefore, we opted to use the distribution of wealth estimated by AAM as our benchmark series. To integrate this distribution in our income data, we use the joint distribution between income and wealth obtained matching IT-SILC and SHIW. First of all, we rank individuals by total net wealth. Thereafter, we associate to each individual the share of wealth obtained by AAM for that specific rank-position (fractile). By multiplying this share by the total National Wealth derived from AAM, we obtain the final individual amount.

This method has the advantage of keeping the same distribution of wealth found by AAM at the fractile level, while allowing us to keep the same relationship between income and wealth derived from the SHIW.

As a further detail on the composition of wealth, we use the components in SHIW to divide the wealth determined by AAM into seven different categories. For each individual, we calculate the share of wealth held in real estate, business, government bonds, equity shares and other securities, valuables, deposits and savings, liabilities in SHIW. Then we partition the net wealth of the AAM data by multiplying the share of the relative type of wealth to the net wealth held.

We use the same method of ranking households by the level of expenditure (fractiles) for

consumption. Then we attribute the consumption derived from the Household and Budget Survey (HBS) of that specific rank (fractiles) to each household. Finally, we share the family consumption among the members in proportion to the individual post-tax disposable income. Also in this case, we exploit the joint distribution derived from merging SHIW and IT-SILC. At the same time, the use of HBS data allows us to be more consistent with the national statistics derived from IStat on consumption. In addition, the HBS is also extremely precise in terms of the type of expenditures and is conducted with a high degree of rigour with repeated interviews at pre-established intervals to improve reliability of the data collected.

A.3 Distribution of taxes

A.3.1 Direct Taxes

Direct taxes in the National Accounts (NA) are grouped under item D5 and are divided according to the payments made by the various institutional sectors. We can regroup the sectors into three main macro aggregates:

1. The household sector made up of producer and consumer households but also of non-profit institutions serving households;

2. The corporate sector, which includes both financial and non-financial companies;

3. The Public Administration sector (in some cases together with the rest of the world sector).

The IStat data for Public Administrations allow us to divide the Direct Taxes into 35 sub-categories received by the public administration. Of these 35 sub-categories, some are paid by households, others by the corporate sector, and some are paid by all institutional sectors. The IStat, however, does not publish the details of the payments made by the various institutional sectors for each sub-category. Therefore we intend to identify for each of the 35 sub-categories to which institutional sector it refers. Once this division is achieved, we

distribute the value of each tax to individuals. Specifically, we have identified the following 25 categories as direct taxes paid by households, and we distribute them to each individual as follows:

- D51A C01 C W0 taxes on the income of individuals or families excluding income from capital are distributed in proportion to the personal income tax present in SILC;
- D51A C02 C W0 additional regional income tax is distributed in proportion to the personal income tax present in SILC;
- D51A C03 C W0 additional municipal income tax is distributed in proportion to the personal income tax present in SILC;
- D51A C04 C W0 withholdings on interest and income from capital - households - are distributed in proportion to equities and shares;
- D51A C05 C W0 local income tax (ilor) - households - is distributed in proportion business income and business assets;
- D51A C06 C W0 tax on income from mobile wealth is distributed in proportion to bonds, equities and shares;
- D51A C07 C W0 complementary and additional taxes are distributed in proportion to financial assets (in AAM, SHIW or SILC);
- D51A C08 C W0 gescal contributions paid by employees are distributed in proportion to the income of employees;
- D51A C09 C W0 tax on the increase in the value of real estate (invim) - households - is distributed in proportion to real estate properties;
- D51A C10 C W0 withholdings on profits distributed by companies - households - are distributed in proportion to equities and shares;

- D51A C11 C W0 municipal tax on industry and the arts and professions (iciap) is distributed in proportion to income from self-employment;

- D51A C12 C W0 taxes on life insurance and supplementary pension are distributed in proportion to deposits, equities and shares;

- D51A C13 C W0 tax on mathematical insurance reserves is distributed in proportion to the deposits, equities and shares;

- D51A C14 C W0 substitute tax on "cedolare secca" real estate rental income is distributed using real estate properties and tax income tables. In particular, the flat-rate income is reported in the IRPEF reports by region or by income class up to 2016. From 2016 onwards, the tables with a double classification of income class and region are available.

By identifying the same income classes and regions in the SILC data, we divide the value of the flat-rate income in that class in proportion to the real estate properties in that class and region. Having thus distributed the income from cedolare-secca, we can calculate the related tax. With this division, we remain consistent with the MEF considerations regarding the regional distribution of the cedolare-secca (Di Caro et al. 2018).

- D51C T C W0 taxes on profits are distributed in proportion to equities and shares;

- D51D C01 C W0 tax, skill games and prediction competitions (direct) is distributed in proportion to the Pre-tax National Income;

- D51E C01 C W0 additional to the state and local income taxes (8%) are distributed in proportion to the Pre-tax National Income;

- D59A C01 C W0 municipal real estate tax (ici) - building areas - is distributed in proportion to real estate properties;

- D59A C02 C W0 municipal real estate tax (building areas) is distributed in proportion to real estate properties;
- D59A C03 C W0 imp. boats and aircraft is distributed in proportion to Valuables assets;
- D59A C04 C W0 television subscription fee for private household use is distributed in proportion to the Pre-tax National Income;
- D59D C01 C W0 tax on driving licenses is distributed in proportion to the Pre-tax National Income;
- D59D C02 C W0 car taxes paid by households are distributed in proportion to the Pre-tax National Income.

We then identify the following eight categories as direct taxes typically paid by businesses distributed through business assets, equities and shares. The variables are:

- D51B C01 C W0 withholdings on interest and income from capital - companies;
- D51B C02 C W0 taxes on corporate income or profits, excluding equity ones;
- D51B C03 C W0 local income tax (ilor) - companies;
- D51B C04 C W0 tax on the increase in the value of properties (invim) - companies;
- D51B C05 C W0 corporate and bond tax;
- D51B C06 C W0 withholdings on profits distributed by companies - firms;
- D51B C07 C W0 tax on corporate equity;
- D51B C08 C W0 new substitute tax revaluation of company assets.

The last direct tax D59F T C W0 is considered as a tax paid by other institutional sectors. It is essential to underline that this division of direct taxes, between the institutional sectors Households, Businesses and other sectors, does not perfectly reflect the division present in the National Accounts. All institutional sectors pay some of the items reported, and it is not always possible to make a precise distinction between the amounts paid by households and the one paid by firms. However, the division carried out allows a good approximation of the amount paid by each institutional sectors. In Table 2 we compare the aggregate of direct taxes D5 D, as published in the National Accounts, and the sum of the 35 sub-categories of direct taxes divided with our identification between families, companies and other sectors. As can be seen, the share paid by households and businesses with our subdivision is always very close to the percentage actually reported in the National Accounts in the aggregates D5.

Table 2: **Conti Nazionali Imposte Dirette**

Year	*Type*	*Total Direct Taxes D5*	*Direct taxes paid by Familes*	*Direct taxes paid by Firms*	*Direct taxes paid by other sectors*
2004	Our division	185,289.00	83.06%	16.26%	0.68%
2004	NA original	185,289.00	80.92%	18.13%	0.95%
2005	Our division	191,001.00	81.84%	17.61%	0.56%
2005	NA original	191,001.00	81.31%	17.73%	0.96%
2006	Our division	213,472.00	78.70%	20.54%	0.76%
2006	NA original	213,472.00	78.44%	20.81%	0.75%
2007	Our division	233,507.00	77.66%	21.65%	0.69%
2007	NA original	233,507.00	77.05%	21.81%	1.15%
2008	Our division	239,880.00	79.39%	19.97%	0.64%
2008	NA original	239,880.00	79.00%	20.14%	0.86%
2009	Our division	222,527.00	82.31%	16.83%	0.86%
2009	NA original	222,527.00	82.28%	17.13%	0.59%
2010	Our division	226,675.00	82.95%	16.22%	0.84%
2010	NA original	226,675.00	83.01%	16.52%	0.46%
2011	Our division	226,939.00	83.13%	15.78%	1.09%
2011	NA original	226,939.00	83.36%	16.21%	0.42%
2012	Our division	239,794.00	83.05%	15.76%	1.19%
2012	NA original	239,794.00	83.10%	16.25%	0.65%
2013	Our division	241,066.00	82.45%	16.77%	0.78%

2013	NA original	241,066.00	82.28%	17.07%	0.65%
2014	Our division	237,175.00	84.30%	14.83%	0.87%
2014	NA original	237,175.00	83.89%	14.98%	1.13%
2015	Our division	242,579.00	85.30%	13.78%	0.92%
2015	NA original	242,579.00	84.95%	14.07%	0.98%

A.4 Social Security Contributions

Social Security Contributions in the National Accounts provided by IStat allow the division between employers and workers, and also between employees (D613CE C) and self-employed workers (D613CNAS C). Therefore, we distribute the contributions from employees (D613CE C) in proportion to the social contributions paid by employees present in the IT-SILC (variable "csdi") and distribute the contributions paid by self-employed workers (D613CNAS C) in proportion to the contributions paid by self-employed workers ("csa" variable). For the contributions paid by employers, we distribute both the actual and figurative contributions of the entire economy, in proportion to the relevant variable in IT-SILC ("csda" variable). It is helpful to notice that the amount of contributions present in SILC is always very close to the totals present in the National Accounts.

A.5 Distributional National Accounts

Following the DINA guidelines, we recreate the four distributions of national income: (i) Factor Income, i.e. the distribution of income derived solely from the remuneration of labour and capital, which therefore includes the contributions paid by workers; (ii) Pre-tax National Income, or the distribution of income deriving from work and capital net of contributions paid but with the addition of transfers from the pension system; (iii) Post-tax Disposable Income, calculated as the Pre-tax National Income but after having subtracted direct and indirect taxes; (iv) Post-tax National Income, calculated as Post-tax Disposable Income but including public expenditure as a source of income of the population. Furthermore, for the household sector, IStat releases the national accounts with a regional breakdown. By exploiting this

information, we therefore recreate the distribution of the household sector not only at the national but also at the regional level.

In order to proceed with the distribution of income variables that are present in the National Accounts, we assume that the net amount declared in the survey is the most truthful information. We take the net variables and add the distribution of taxes paid adjusted by National Accounts as described in Appendix A.3. Thus we obtain a new provisional value for gross incomes given by net incomes in SILC plus the distribution of taxes. We use this provisional gross income to distribute the income components of the National Accounts. Moreover, suppose the total of this provisional gross income (i.e. net income in SILC plus the distribution of taxes) is higher than the gross income reported in the NA. In that case, after distributing the values of the National Accounts in proportion to the provisional gross incomes, we create new values of net income given by the difference between the final adjusted gross incomes and the distribution of taxes. Using this method, we no longer have to change the distribution of taxes paid, but we instead have a new net income consistent with the distribution of gross income and taxes.

A.5.1 Factor Income

Factor Income is income that remunerates self-employed workers, employees and investments, with their sum being equal to the total net national income in the National Accounts. For the household sector, these incomes are the following:

- D11 C gross wages net of contributions to be paid by employees D613CE C, that we distribute in proportion to the income from employees;
- D12 C social contributions paid by employers, distributed in proportion to the contributions paid by employers;
- B2N net operating profit. As reported in SNA08, this is the category of income identifying imputed rents in the household sector. Therefore, we distribute it based on real

estate properties;

- B3N net mixed income of contributions to be paid by self-employed workers D613CNAS C. According to IStat[5] this income component is the sum of self-employment income and rents from real estates of the household sector. Therefore, we split this income source using the same ratio of self-employment income and income from rents present in the IT-SILC. Then we distribute the portion relative to rents in proportion to the real estate properties and the rest in proportion to self-employment income;

- D41 C-D interest, being the total interest income received. We distribute them in proportion to bonds, deposits and savings, equities and shares;

- D42 C-D profits distributed by companies. We distribute them in proportion to business assets, equities and shares;

- D44 C-D other investment income. We distribute them in proportion to deposits and savings, equities and shares;

- D45 C-D rental of land and exploitation rights of fields. We distribute them in proportion to the capital income previously distributed.

The sum of these incomes, adding the previously distributed social contributions, perfectly recreates the net national income of the institutional sector of households. This method of distribution of capital income is equivalent to that made by Piketty, Saez, and Zucman (2018) for the United States, in which it is assumed that the return on capital is constant for each level of wealth. This is a relatively strong assumption, as recent findings pointed out that a higher level of wealth is associated with a higher rate of returns (Fagereng et al. 2020; Iacono and Palagi 2021; Bach, Calvet, and Sodini 2020).

To this part of national income, we also want to add the incomes of the corporate sector. These incomes, also called retained earnings, can be considered as income for the business

[5] https://www.istat.it/it/files//2020/12/REPORT-CONTI-TERRITORIALI_2019.pdf

owners, even if it is not distributed yet. Therefore, we also distribute the net income of the corporate sector (B5N) in proportion to equities and shares. As for the income obtained by the public administration sector, in line with the DINA guidelines, it is distributed in proportion to the Factor-Income, so that only the absolute level of income earned is affected and not their relative shares.

A.5.2 Pre-tax National Income

To construct the Pre-tax National Income distribution, we must start from the distribution of the Factor-Income. However, it is necessary to include the transfers for the pensions due for the payment of contributions in working age and deduct the social contributions paid. In the National Accounts, pensions and transfers are reported in aggregate D62. This, in turn, is composed by the following sub-items:

- D621 and D622, i.e. pensions due to the payment of social contributions;
- D623, i.e. social transfers not subject to the payment of contributions.

However, IStat does not release the composition of aggregate D62. In order to make this division, we follow BCG and use OECD data on social expenditure to identify component D623 and subtract it from aggregate D62 and obtain the total of the pensions paid for contributory rights.

Furthermore, to obtain the total net national income, we must also consider the difference between the expenditure on social contributions of the corporate sector and the public sector. In fact, some companies set up their own social security systems in which families can participate by paying a periodic fee. These companies generally have a surplus account and therefore have more contribution income than the transfers they issue. This surplus between expenditure and contributions must be included in the net income of firms.

The opposite is true for public administrations. In fact, they usually issue more contributions than what they receive, so we must include this debt in the net income of the general

government. In addition, to perfectly reach the total national income, we add to the public sector any debts or credits (albeit small) existing with the sector of the rest of the world and with households.

A.5.3 Post-tax Disposable and National Income

The Post-tax disposable income aims at reconstructing the post-tax and post-transfer income received by the population, excluding in-kind transfers. Therefore, we deduct all direct and indirect taxes that we distributed as previously described, while adding back subsidies on production and distributing social assistance benefits in cash. This income concept is lower than the National Income by construction since collective public expenditure is missing from the totals. To revert to the total National Income, we use the income concept of the Post-tax National income, in which we add back the public expenditure and any deficit/surplus of the government sector. We follow the DINA guidelines distributing collective expenditure as an homogeneous lump sum for all the population and distribute the rest in proportion to the post-tax disposable income. It is important to notice that any distributional choice of public expenditure has crucial consequences in terms of inequality statistics but adding back public expenditure has the advantage to make comparable countries with highly different public spending. Since Italy has a very fragmented public expenditure that vastly depends on the region of residence, we also depart from the standard distributional assumption of the DINA guidelines. By relying on data on public expenditure by regions published by the "Ragioneria Generale dello Stato", we first distribute the national public expenditure to regions and, then, we distribute the spending among the residences of the regions. However, using a regional distribution rather then the national total, does not significantly change the distributional impact.