WID.world – Working Paper N° 2016/2

Distributional National Accounts Guidelines: Methods and Concepts Used in WID.world

Facundo Alvaredo
Anthony B. Atkinson[Ψ]
Lucas Chancel
Thomas Piketty
Emmanuel Saez
Gabriel Zucman

First version: December 12th, 2016
This version: April 5th, 2018

WID.world Working Paper

Distributional National Accounts (DINA) Guidelines:

Concepts and Methods used in WID.world *

(World Inequality Database, WID.world)

Facundo Alvaredo, Anthony B. Atkinson[Ψ], Lucas Chancel,

Thomas Piketty, Emmanuel Saez, Gabriel Zucman

First version: December 12th, 2016

This version: April 5th, 2018

(An updated and extended version of the guidelines was published in 2020 and is available on WID.world)

* These Guidelines aim to synthesize the concepts and methods used in *WID.world* (*World Inequality Database*) (http://WID.world). They are subject to revision and will be updated on-line.

[Ψ]Honorary member.

Section 1. Introduction

The purpose of these DINA Guidelines is to present the concepts, data sources and methods used in the World Inequality Database (WID.world, http://WID.world). These Guidelines are subject to revision and will be regularly updated on-line. Before we describe the organization of these Guidelines, it is useful to start with a brief history of WID.world.

During the past fifteen years, the renewed interest for the long-run evolution of the distribution of income and wealth gave rise to a flourishing literature. In particular, by combining historical fiscal and national accounts data in a systematic manner, a succession of studies has constructed top income share series for a large number of countries (see Piketty 2001, 2003, Piketty-Saez 2003, and the two multi-country volumes on top incomes edited by Atkinson-Piketty 2007, 2010; see also Atkinson-Piketty-Saez 2011 and Alvaredo-Atkinson-Piketty-Saez 2013 for surveys of this literature). These projects generated a large volume of data, intended as a research resource for further analysis, as well as a source to inform the public debate on income inequality. To a large extent, this literature follows the pioneering work and methodology of Kuznets (1953) and Atkinson and Harrison (1978) on the long-run evolution of income and wealth distribution, and extends it to many more countries and years.

The *WTID* (*World Top Incomes Database*) was created in January 2011 in order to provide easy on-line access to all series. It currently includes homogenous series on income inequality for more than 30 countries, spanning over most of the 20th and early

21st centuries, while over 40 additional countries are under study. More than 100 researchers from all parts of the world have contributed to the *WTID*. The key novelty has been to exploit fiscal, survey and national accounts data in a systematic manner. This allowed us to compute longer and more reliable top income shares series than previous inequality databases (which generally rely on self-reported survey data, with large under-reporting problems at the top, and limited time span). These series had a large impact on the global inequality debate.

In December 2015, the *WTID* was subsumed into the *World Inequality Database* (http://WID.world). In addition to the WTID top income shares series, this first version of *WID* included an extended version of the historical database on the long-run evolution of aggregate wealth-income ratios and the changing structure of national wealth and national income first developed by Piketty-Zucman 2014 (see also Piketty, 2014, for an attempt to propose an interpretative historical synthesis on the basis of this new material and of the top income shares series). We changed the name of the database from WTID to WID in order to express the extension in scope and ambition of the database and the new emphasis on both wealth and income.

In conjunction with the development of a novel website with new data visualization possibilities (the new website was made public for the first time on http://WID.world in January 2017), the WID.world project is currently involved in major extensions in three directions, which will be gradually implemented. First, we pursue our efforts to cover more and more countries, in particular among the emerging countries of Asia, Africa and Latin America. Next, we plan to provide more and more series on wealth-income ratios and the distribution of wealth, and not only on income. Finally, we aim to offer

series on the entire distribution of income and wealth, from the bottom to the top (and not only for top shares). The overall objective is to be able to produce *Distributional National Accounts (DINA)*, that is, to provide annual estimates of the distribution of income and wealth using concepts of income and wealth that are consistent with the macroeconomic national accounts. This also includes the production of synthetic income and wealth micro-files, which will also be made available online. Such data can play a critical role in the public debate, and can be used as a resource for further analysis by various actors of the civil society and the academic, business and political community. The long-run aim is to release synthetic income and wealth *DINA* micro-files for all countries on an annual basis.

It is worth stressing that the new WID.world database has both a macro and a micro dimension. Our objective is to release homogenous series both on the macro-level structure of national income and national wealth, and on the micro-level distribution of income and wealth, using consistent concepts and methods. By doing so, we hope to contribute to reconcile inequality measurement and national accounting, i.e. the micro-level measurement of economic and social welfare and the macro-level measurement. In some cases this may require to revise key national accounts concepts and estimates. By combining the macro and micro dimensions of economic measurement, we are of course following a very long tradition. In particular, it is worth recalling that Kuznets was both one of the founders of U.S. national accounts and the author of the first national income series, and also the first scholar to combine national income series and income tax data in order to estimate the evolution of the share of total income going to top fractiles in the U.S. over the 1913-1948 period (see Kuznets 1953). We are simply pushing this effort further by trying to cover many more countries and years,

and by studying wealth and its distribution and not only income (a line of research pioneered by Atkinson and Harrison 1978, who combined historical inheritance tax data with capital income data and wealth surveys to study the long-run evolution of wealth distribution for Britain over the 1922-1972 period).

Needless to say, such an ambitious long-term objective - annual distributional national accounts for both income and wealth and for all countries in the world - will require a very broad international and institutional partnership. We certainly do not claim that the WID.world project in its current form has the capability to achieve this objective alone. The WID.world project started as an informal academic network, and it is now financed by a number of research grants by public research agencies - including the European Research Council - and non-profit institutions (more on this on-line). It will keep evolving in the future, and in order to achieve its long-run objective new partnerships will undoubtedly need to be developed, in particular with international organizations and statistical agencies. Our work should be viewed as one step in a long, collective and cumulative research process.

As the *WID*.world project is expanding in scale and ambition, we believe that it is time to further clarify and homogenize its concepts and methods. The purpose of these *DINA Guidelines* is to present the concepts and methods that will be followed in the database. These guidelines are provisional and subject to revisions. Additional details are provided in the research papers developing prototype *DINA* estimates for specific countries (see in particular Piketty-Saez-Zucman 2016 for the U.S., Garbinti-Goupille-Piketty 2016, 2017 and Bozio-Garbinti-Goupille-Piketty 2017 for France, Alvaredo-Atkinson-Morelli 2017 for the U.K., Piketty-Yang-Zucman 2017 for China). The purpose

of these *DINA Guidelines* is to synthesize the lessons from these country-specific works and provide guidance for future countries. The *Guidelines* will be updated accordingly as more countries become available.

We should stress at the onset that our methods and series are and will always be imperfect, fragile and subject to revision. We attempt to combine the different data sources that are available (in particular fiscal data, survey data and national accounts) in a more systematic way than what was done before. We also try to provide a very detailed and explicit description of our methodology and sources, so that other users can contribute to improving them. But our series and methods will always be imperfect and should be viewed in the perspective of a long, cumulative, collective process of data construction and diffusion.

The concepts and methods used in *WTID* series were initially exposed in the two collective volumes edited by Atkinson-Piketty (2007, 2010) and in the corresponding country chapters and research articles. In principle, all series follow the same general methods: following the pioneering work of Kuznets (1953), they combine income tax data, national accounts, and Pareto interpolation techniques in order to estimate the share of total income going to top income groups (typically the top decile and the top percentile). However, despite our best efforts, the units of observation, the income concepts, and also the Pareto interpolation techniques, were never made fully homogenous over time and across countries. Moreover, for the most part we restrict our attention to the top decile income share, rather than the entire distribution of income and wealth.

In contrast, the *DINA* series and associated synthetic micro-files aim to be fully homogenous across all of these dimensions (or at least to make much more explicit the remaining heterogeneity in data construction), and most importantly to provide more detailed and comprehensive measures of inequality. In *DINA* series, inequality is always measured using homogenous observation units, and taxable income reported on fiscal returns is systematically corrected and upgraded in order to match national accounts totals separately for each income categories (wages, dividends, etc.), using various sources and imputations methods. We address each of these issues below, as well as a number of new issues related to the fact that we now aim to produce series on wealth (and not only on income) and on the entire distribution (and not only on top shares). The two main data sources used in *DINA* series continue to be income tax data and national accounts (just like in the *WTID* series), but we use these two core data sources in a more systematic and consistent manner, with fully harmonized definitions and methods, and together with other sources such as household income and wealth surveys, inheritance and wealth tax data, as well as wealth rankings provided by "rich lists" compiled by the press. In most cases, the general trends in inequality depicted in the *WTID* series will not necessarily be very different in *DINA* series. However the latter will allow for more precise comparisons over time and across countries, more systematic world coverage, and more consistent analysis of the underlying mechanisms.[1]

The *DINA Guidelines* are organized as follows. Section 2 discusses units of observation (from individual level to the world level) and inequality measures used in the *WID* (from bottom percentiles to top percentiles). Section 3 presents the income

[1] As new *DINA* series become available, we will systematically compare the inequality trends obtained in the old and the new series and analyze the sources of biases.

concepts: pre-tax national income, pre-tax factor income, post-tax disposable income and post-tax national income. Section 4 presents the wealth concepts (personal wealth, private wealth, public wealth and national wealth), as well as the corresponding notions of capital income flows and rates of return that are used in the *WID*. Section 5 presents the basic imputations methods that we use in order to reconcile income tax returns micro files with national accounts. Section 6 discusses the methods used to reconcile the different data sources on wealth inequality. Section 7 addresses the case of countries and years with limited income and wealth data: typically, income tax tabulations instead of income tax micro files, and/or limited aggregate wealth estimates and billionaire lists instead of detailed balance sheets and distributional data. For many developing countries (and also for a number of developed countries with limited access to fiscal, national accounts and wealth data), this is to a large extent the most important section for the years to come. Section 8 concludes by listing a number of pending issues.

Section 2. Units of observation

Section 2.1. Micro-level observation units: equal-split adults and individualistic adults

One of the major limitations of the *WTID* series so far was the lack of homogeneity of the micro-level observation unit. Most *WTID* series were constructed by using the "tax unit" (as defined by the tax law of the country at any given point in time) as the observation unit. In joint-taxation countries like France or the U.S., the tax unit has always been defined as the married couple (for married individuals) or the single adult (for unmarried individuals), and the top income shares series that were produced for these two countries (see Piketty, 2001, 2003, and Piketty and Saez, 2003) do not include any correction for the changing structure of tax units (i.e. the combined income of married couples is not divided by two, so couples appear artificially richer than non-married individuals).[2] This is problematic, since variations in the share of single individuals in the population, or in the extent of assortative mating in couples, could potentially bias the evolution of income inequality in various and contradictory ways. In some other countries, the tax system switched to individual taxation over the course of the history of the income tax (e.g., in 1990 in the U.K.), which creates other comparability problems in the *WTID* series (see Atkinson, 2005, 2007).

In order to correct for these biases, our *DINA* series attempt to use homogenous observation units. Generally speaking, our benchmark unit of observation is the adult individual. That is, our primary objective is to provide estimates of the distribution of

[2] I.e. the top 10% income share in WTID series relates to the income share going to the top 10% tax units with the highest incomes (irrespective of the size of tax units, which means that married couples with two earners are likely to be over-represented at the top of the distribution).

income and wealth between all individuals aged 20-year-old and over (such as the shares of income and wealth going to the different percentiles of the distributions of income and wealth). Whenever possible, we also aim to construct estimates of individual income and wealth distribution that can be decomposed by age, gender and numbers of dependent children. Ideally, we aim at producing synthetic micro-files providing the best possible estimates of the joint distribution of age, gender, numbers of dependent children, income and wealth between adult individuals. But at the very least we want to be able to describe the distribution of income and wealth between all adult individuals.

One key question is how to split income and wealth between adults who belong to a couple (married or not) and/or to the same household (i.e. adults who live in the same housing unit). To the extent possible, we aim to produce for each country two sets of inequality series: “equal-split-adults series” and “individualistic-adults series”. In the equal-split series, we split income and wealth equally between adults who belong to the same couple (and/or the same household; more on this below). In the individualistic series, we attribute income and wealth to each individual income earner and wealth owner (to the extent possible; more on this below).

We should make clear that both series are equally valuable in our view. They offer two interesting and complementary perspectives on different dimensions of inequality. The equal-split perspective assumes that couples redistribute income and wealth equally between its members. This is arguably a very optimistic and/or naïve perspective on what couples actually do: bargaining power is typically very unequal within couples, partly because the two members come with unequal income flows or wealth stock. But

the opposite perspective (zero sharing of resources) is not realistic either, and tends to underestimate the resources available to non-working spouses (and therefore to overestimate inequality in societies with low female participation to the labor market). By offering the two sets of series, we give the possibility to compare the levels and evolutions of inequality over time and between countries under these two different perspectives. Ideally, the best solution would be to organize synthetic micro files in such a manner that the data users can compute their own inequality series based upon some alternative sharing rules (e.g. assuming that a given fraction of the combined income of couples is equally split) and/or some alternative equivalence scales (e.g. dividing the income of couples by a factor less than two). This is our long-run objective.

Regarding the equal-split series, an important question is whether we should split income and wealth within the couple (narrow equal-split) or within the household (broad equal-split). In countries with significant multi-generational cohabitation (e.g. grand-parents living with their adult children), this can make a significant difference (typically broad equal-split series assume more private redistribution and display less inequality). In countries where nuclear families are prevalent, this makes relatively little difference. Ideally both series should be offered. We tend to favor the narrow equal-split series as benchmark series, both for data availability reasons (fiscal data is usually available at the tax unit level, which in a number of countries means the married couple or the non-married adult) and because there is possibly more splitting of resources at the narrow level (which is also arguably the reason why fiscal legislation usually offers the possibility of joint filling and taxation at the level of the married couple rather than at the level of the broader household, whose exact composition can vary and is not

regulated by a formal legal relationship).[3] However in countries where fiscal sources are limited and where we mostly rely on household survey data (e.g. in China), it is sometime easier to compute the broad equal-split series. This should be kept in mind when making comparisons between countries (see e.g. the discussion in Piketty, Yang and Zucman (2017) and the comparison between DINA series for China, France and the United States).

Finally, when we look at the inequality of post-tax disposable income, we also introduce dependent children into the analysis, in order to be able to compute the relevant cash and in-kind transfers to the parents (family benefits and tax credits, education spending, and so on; see the discussion in section 6 below).

In the individualistic series, observed labor income and pension income is attributed to each individual recipient. This is easy to do in individual-taxation countries like the U.K. today, where by definition we observe incomes at the individual level. In general, labor income and pension income are also reported separately for each spouse in the tax returns and income declarations used in joint-taxation countries like France. In some cases, however, e.g. in U.S. public-use tax files, we only observe the total labor or pension income reported by both spouses, in which case we need to use other sources and imputations techniques in order to split income appropriately between spouses (see Piketty-Saez-Zucman 2016).

[3] We usually include civic unions (PACS in France, etc.) in married couples, to the extent that they are treated in the same way as married couples by fiscal legislation. See discussion in specific country papers.

Issues are more complicated for capital income flows. In individual-taxation countries, we usually observe capital income at the individual level, so there is no particular difficulty. However in joint-taxation countries, capital income is usually not reported separately for both spouses, and we generally do not have enough information about the marriage contract or property arrangements within married couples to be able to split capital income and assets into community assets and own assets. So in joint-taxation countries we simply assume in our benchmark series that each spouse owns 50% of the wealth of a married couple and receives 50% of the corresponding capital income flow. If and when adequate data sources become available, we might be able to offer a more sophisticated treatment of this important issue.[4]

2.2. Aggregate observation unit (country, regions, world), g-percentiles, micro-files

Our basic objective in constructing *DINA* series is to present the best possible estimates of the distribution of income and wealth between all adult individuals living in a given country during a given year. However, we also want to be able to measure inequality for different geographical units than the country level, e.g. in some cases at the sub-national level (regions of given country),[5] as well as at the continental level (regions of the world, such as Europe) or at the world level.

[4] In order to be consistent, we also allocate to each spouse 50% of the estimated capital share of mixed (self-employment) income; in contrast, we allocate 100% of the estimated labor share of mixed income to the self-employed adult individual himself or herself (see section 4 below on how we spit mixed income into labor and capital components). Note that we also split 50-50 the capital income of couples with "civil union contracts" (such as PACS in France), who according to French law also fill joint returns and report a single capital income amount (just like for married couples).

[5] In some cases, we might indeed be able to provide estimates of the distribution of income and wealth at the sub-national level, e.g. for U.S. states or for major cities. Data on inequality at the sub-national levels are important to better understand the causes and consequences of rising inequality.

This is one of the key reasons why we aim to produce synthetic *DINA* micro-files on the individual-level distribution of income and wealth: such files can be easily aggregated from the country or regional level to the continental or world level. One simply needs to merge the different files, using adequate population weights. In contrast, the *WTID* series usually take the form of top income shares series, typically with thresholds and averages for the top 10% incomes, the top 5%, the top 1%, and so on, which cannot be easily aggregated.[6]

Another key advantage of micro-files is that they will allow us to provide country-level series on thresholds and averages for income and wealth at each percentile of the distribution (together with a finer decomposition within the top percentile). We will indeed provide such a representation of the data on the *WID* website, which for instance can be used to allow individual users to locate themselves easily within the distribution.[7] In addition, micro-files can be used as a resource for further analysis by various actors from academia and the civil society, for instance in order to simulate tax reforms.[8]

At the very least (i.e., for countries/years with very limited data, typically with income tax tabulations instead of micro-files, and no other source of information on age and gender profiles), we aim to produce for each country/year a synthetic micro-file describing the distribution of income and wealth among all adult individuals. Whenever

[6] For a recent attempt to combine household survey data and top income shares series in order to study the recent evolution of income inequality at the world level, see Lakner and Milanovic (2013). One of our objectives is to be able to pursue this kind of approach in a systematic manner.
[7] See for instance the platforms for income distribution and wealth distribution developed by Landais, Piketty and Saez (2011) (see www.revolution-fiscale.fr). We plan to offer a similar platform for all countries and years on the *WID.world* website.
[8] Indeed one of the main motivations behind the prototype DINA microfiles developed for France by Landais-Piketty-Saez (2011) was the provision of an on-line tax reform simulator.

possible, we aim to produce for each country/year a synthetic micro-file describing the joint distribution of age, gender, income and wealth among all adult individuals.

WID.world data and micro-files are made available in two different forms: first by using generalized-percentiles (or g-percentiles) files; next by using large files with representative numbers of synthetic observations (e.g. one million or ten million or more, depending on the size of the country and the needs of the data user).

G-percentiles files use 127 rows: 99 for the bottom 99 percentiles, 9 for the bottom 9 tenth-of-percentiles of the top percentile, 9 for the bottom 9 one-hundredth-of-percentiles of top tenth-of-percentile, and 10 for the 10 one-thousandth-of-percentile of the top one-hundredth-of-percentile. Files at the g-percentile level include for each g-percentile row the average income and the corresponding income threshold (see appendix table A1). These g-percentile files are sufficient for most users, e.g. they allow to compute percentile shares and synthetic inequality indexes such as Gini coefficients.

Large files can be generated by the data user by specifying the number of synthetic observations in the “generalized Pareto interpolation” (gpinter) web interface available in the methodology section of the WID website (http://WID.world/gpinter/). The interface then generates a synthetic file with the required number of observations, using the generalized Pareto curves interpolation techniques developed by Blanchet-Fournier-Piketty 2017 (see section 7 below for a brief description). The interface also allows the users to merge income and wealth distributions for any given of country,

e.g. to compute the g-percentiles of the combined country from the g-percentiles of each individual country (or region).

Section 3. Income concepts

Section 3.1. Reconciling inequality measurement and national accounts

One of the other major limitations of the *WTID* series (together with the observation unit problem referred to in section 2, and together with the fact that most existing series focus upon pre-tax inequality and largely ignore post-tax inequality) is the lack of homogeneity of the income concept. Most *WTID* series were constructed by using some kind of "fiscal income" concept, i.e., total income that is or should be reported on income tax declarations (before any specific deduction allowed by fiscal legislation).[9] The problem is that such concepts naturally vary with the tax system and legislation that is being applied in the country/year under consideration. It is worth stressing that we did not attempt until now to correct in a systematic manner for the fact that some forms of income (e.g. a number of specific components of capital income) that are legally not subject to tax and do not appear on income tax declarations.[10] As a consequence, the "fiscal income" concept used in *WTID* varies over time and across countries, which in some cases might create biases.[11]

[9] ”Fiscal income” is broader and somewhat more homogenous than “taxable income”, which we define as fiscal income minus existing income tax deductions (which typically vary a lot across countries and over time with the tax legislation). For instance, in France, all wage earners benefit from a 10% standard deduction for "professional expenses" (up to ceiling). In the case of France, like in most countries, the raw tax data generally use the concept of “taxable income” (post-deductions income), and a number of corrections were applied so that *WTID* series refer to “fiscal income” (pre-deductions income). Although the “fiscal income” concept in WTID series is broader than “taxable income”, it is not sufficiently broad and homogenous over time and across countries.

[10] Sometime some forms of income are not taxable but are reported on tax returns, in which case we usually include them in the "fiscal income" concept used in *WTID* series.

[11] In order to limit biases, we always attempt to use the same "fiscal income" concept for the numerator and the denominator in *WTID* series. But this is clearly not sufficient, especially given that we observe in many countries a tendency for more and more components of capital income flows to be exempt from the progressive income tax base and often to disappear from income tax declarations and statistics all together. As new *DINA* series become available, we will systematically compare the inequality trends obtained in the old and the new series and analyze the sources of biases.

In contrast, the income concepts that we use in *DINA* series are defined in the same manner in all countries and time periods, and aim to be independent from the fiscal legislation of the given country/year. As we explain below, the four basic pre-tax and post-tax income concepts that we use to measure income inequality are anchored upon the notion of "national income" (i.e. gross domestic product, minus consumption of fixed capital, plus net foreign income) and are defined by using the same concepts as those proposed in the latest international guidelines on macroeconomic national accounts, as set forth by the 2008 UN System of National Accounts (SNA) (see U.N. National Accounts website and SNA 2008 online guideline page and SNA 2008 pdf guideline). In what follows and in our on-line database, we often refer to the classification codes from *SNA 2008* or from the European System of Accounts (*ESA 2010*).[12] In some countries, and/or for some earlier years, available national accounts series still follow the earlier system of international guidelines, namely *SNA 1993* (or the European version, *ESA 1995*). The differences between the two systems are usually minor; in the few cases where there are significant differences we mention them below or in the country-specific papers.[13]

We should make clear at the onset that our choice of using national accounts income and wealth concepts for distributional analysis certainly does not mean that we believe that these concepts are perfectly satisfactory or appropriate. Quite the contrary: our

[12] *ESA 2010* is the European Union implementation of *SNA 2008*; both systems are virtually identical (see Eurostat National Accounts website, *ESA 2010 online guideline page* and *ESA 2010 pdf guideline*). Note that the *ESA 2010* classifications sometime provide more detailed subcategories than *SNA 2008* classifications, e.g. regarding non-financial assets (see section 4 below). The classifications used by U.S. national accounts are somewhat different, and whenever necessary we reclassify them in order to match the international classifications (see Piketty, Saez and Zucman (2016)).

[13] The main innovation between SNA 1993/ESA 1995 and SNA 2008/ESA 2010 is the fact that research and development is now explicitly treated as investment and capital accumulation (with the introduction of a new non-financial asset category: AN117, "Intellectual property product"). See this Eurostat Manual describing the main changes between the two systems.

view is that official national accounts statistics are insufficient and need to be greatly improved. In particular, one of the central limitations of official GDP accounting is that it does not provide any information about the extent to which the different social groups benefit from growth. By using national accounts concepts and producing distributional series based upon these concepts, we hope we can contribute to address one of important shortcomings of existing national accounts and to close the gap between inequality measurement and national accounts, and also maybe between the popular individual-level perception of economic growth and its macroeconomic measurement.

The other reason for using national accounts concepts is simply that these concepts represent at this stage the only existing systematic attempt to define notions such as income and wealth in a common way, which (at least in principle) can be applied to all countries and that is independent from country-specific and time-specific legislation and data sources. These concepts need to be refined, but in order to do so and to propose amendments and improvements, we feel that the best way to proceed is to start from them, use them and modify them when needed. The alternative would be start from scratch and propose entirely new definitions of income, output and wealth, which does not seem realistic nor desirable.

Whenever our WID.world aggregate national income series depart from official series, we will make it explicit and justify our choices (and in some cases make suggestions for changes to future SNA definitions). For instance, we aim to correct official series on foreign capital income flows in order to take into account offshore wealth (using estimates of offshore wealth and its geographical distribution recently proposed by Zucman, 2013, 2014) and to ensure that these flows sum up to zero at the world level.

The resulting adjustments to aggregate national income should be viewed as provisional and are relatively small for most countries, but at least this allows us to present global series on income and wealth that are logically consistent, and to open the way for more systematic measurement effort in this direction (see Blanchet and Chancel, 2016, for a description of our methodology and results). In some cases the inclusion of offshore wealth can make a large difference, both at the aggregate and the distributional levels (e.g. in Russia and Gulf countries the share of financial wealth held offshore seems to exceed 50%; see Zucman 2014, table 1).

Another important limitation of existing official national accounts is the fact that consumption of fixed capital does not usually include the consumption of natural resources. In other words, official statistics tend to overestimate both the levels and the growth rates of national income, which in some cases could be much lower than those obtained for gross domestic product. In the future, we plan to gradually introduce such adjustments to the aggregate national income series provided in the WID.world database. This is likely to introduce significant changes both at the aggregate and distributional level.[14]

We should also make clear that official national accounts are often fairly rudimentary in a number of developing countries (and also sometime in developed countries). Sometime they do not include the level of detail that we need to use the income and wealth definitions proposed below. In particular, proper series on consumption of fixed

[14] A closely related question is the interplay between the global distribution of income and wealth and the global distribution of carbon emissions, an issue which the WID.world database could be used to address in the future. For a preliminary and exploratory attempt to estimate the global individual-level distribution of carbon emissions, see Chancel and Piketty, 2015.

capital and net foreign income are missing in a number of countries, so that official series do not always allow to compute national income. We include in the WID.world database estimates of aggregate and average national income for all countries in the world (including countries for which do not have satisfactory distribution series yet), using a consistent and homogenous methodology (see Blanchet and Chancel, 2016). These estimates should be viewed as provisional and subject to revision. In countries where national accounts are too fragile and where other data sources allow to estimate income and wealth series that are more satisfactory and consistent, we recommend using these other data sources, and we will update our series accordingly. Again, we do not pretend that the concepts and estimates we provide are perfectly satisfactory: our main value added is to be fully explicit about the methods we use to combine the various data sources (which is not always the case for official national accounts and alternative inequality data sets).

Section 3.2. Pre-tax and post-tax income concepts: general definitions

We aim to provide income distribution estimates using four broad concepts of income: pre-tax national income, pre-tax factor income, post-tax disposable income, and post-tax national income. The key difference between pre-tax national income and pre-tax factor income is the treatment of pensions (and other social benefits), which are counted on a distribution basis for pre-tax national income and on a contribution basis for pre-tax factor income (more on this below). We tend to favor the "pre-tax national income" concept, and we view our "pre-tax national income" inequality series as our benchmark series for pre-tax inequality. But we stress that the "pre-tax factor income" inequality series also provide useful and complementary information. Our series are

constructed so that aggregate pre-tax national income and aggregate pre-tax factor income are both exactly equal to aggregate national income (the two distributions vary, but not the aggregate amounts; see below).

Our "post-tax disposable income" series aim to describe post-tax, post-transfer inequality (excluding in-kind transfers such as health and education and other public spending, so that aggregate post-tax disposable income can be substantially less than aggregate national income, typically around 70% of national income in countries where in-kind transfers and public spending represent about 30% of national income). Our “post-tax national income” series include all in-kind transfers and public spending (using various procedures for imputation to individuals, see below), so that aggregate post-tax national income is equal to aggregate national income.

As we shall see, aggregate pre-tax national income, pre-tax factor income, and post-tax national income are all equal to aggregate national income, as defined by *SNA 2008*, but they correspond to different decompositions by income subcomponents and different distributions among individuals. They can be used to analyze the redistributive impact of government taxes, transfers, and spending on a fully comprehensive basis. The various micro-level sources (in particular income tax micro-files and household surveys) and methods that we use to measure and impute these different income components at the individual level will be described in section 5 (and the subsequent sections), and some readers may want go directly to section 5. In the rest of this section, we provide the detailed definitions and decompositions of our four income concepts, using national accounts concepts and guidelines. In order to do so it is useful to start by describing the basic decomposition of national income according to *SNA*

2008. We will then move to pre-tax factor income, pre-tax national income, post-tax disposable income, and post-tax national income, as defined in *DINA* series.

Section 3.3. National income and its decomposition

According to *SNA 2008* (as well as in previous national accounts systems), national income can be defined using either a production approach, or an income approach. By construction, both are fully equivalent (see table 1 and table 2). Note that tables 1 and 2, as well as all subsequent tables presented below, are constructed using the "Sequence of accounts" excel tables provided in *SNA 2008* guidelines.[15] We recommend that readers have a look at the *DINA concepts* excel file (see Appendix to these Guidelines) where we provide formulas relating these tables to the "Sequence of accounts" excel file and to the SNA 2008 classification codes. The actual amounts reported in the "Sequence of accounts" excel tables do not refer to any real country, but the overall structure is broadly representative of the national accounts of advanced economies (we express all amounts in percentage of net national income, together with the raw amounts).

According to the production approach (see table 1), national income is defined as the sum of net domestic product (i.e. gross domestic product, minus consumption of fixed capital) and foreign income (net foreign inflow of capital and labor income). Net domestic product can itself be broken down as the sum of the net value added of each institutional sector (household sector, financial and non-financial corporate sector, government sector, non-profit sector) and of "taxes on products" (i.e. value-added type

[15] The original "Sequence of accounts" excel table published by *SNA 2008* can be found here, and the description of the "Sequence of accounts" can be found in annex 2 of SNA 2008 guidelines.

taxes and other product taxes, which according to *SNA 2008* are not attributed to the value-added of any particular sector).

According to the income approach (see table 2), national income is defined as the sum of primary incomes of each institutional sector. Primary income of the household sector (including unincorporated businesses) is by far the largest component (83% of national income according to the example chosen in the SNA 2008 Sequence of accounts), and is equal to the addition of total employee compensation (including all employer social contributions), net mixed income (i.e. self-employment income), net operating surplus of the household sector (i.e. rental value of housing owned by households, whether it is owner-occupied or rented to other households),[16] and net property income received by households ("property income", as defined by *SNA 2008*, can be further decomposed into interest, dividends, etc., and other financial income flows; we return to this decomposition of capital income flows in section 4).[17]

Primary income of the corporate sector (6% of national income according to the example chosen in the SNA 2008 Sequence of accounts; see table 2) corresponds to

[16] To be precise, the net operating surplus of the household sector is equal to the net operating surplus of the household housing sector. Three remarks are in order. First, household housing stock excludes pre-tax the stock of housing owned by nonprofits, corporations and the government. Second, the net operating surplus of the household housing sector is net of any intermediate consumption, including consumption of financial services indirectly measured (FISIM) supplied by mortgage providers. Because there is substantial cross-country heterogeneity in the way FISIM are measured, comparisons of housing products across countries are rendered somewhat difficult. Whenever necessary and possible, this should be corrected and homogenized (see country specific studies). Third, the net operating surplus of the household housing sector is equal to housing rents (net of intermediate consumption, but gross of mortgage interest payments) plus a small flow of current transfers, typically insurance payments.

[17] Note that "property income" (D4), as defined by to *SNA 2008*, also includes a non-financial income flow, namely "rent" (D45), which by definition excludes housing rents and solely includes rent on natural resources (cultivated land, subsoil assets, etc.). On table 2 and subsequent tables, we choose to include the corresponding net flow received by households with net mixed income rather than with other property income flows (so as to be consistent with the asset categories that we use on tables 6-8; see section 5 below).

undistributed profits (before deduction of the corporate tax): it is equal to the net operating surplus of non-financial and financial corporations, plus the property income that they receive from themselves and other sectors, minus the property income that they pay to themselves and other sectors.

According to *SNA 2008*, primary income of the household and corporate sectors is computed before deduction of direct taxes (in particular before deduction of personal and corporate income taxes), but after deduction of "taxes on production" (D2), which are defined as the sum of "taxes on products" (D21, including value-added type taxes and other product taxes) and "other taxes on production" (D29, including a large number of various taxes such as property taxes on housing, land or buildings used by households or corporations).

Primary income of the government sector (including all public administrations and government agencies, at the national, regional and local levels; about 10% of national income according to the example chosen in the SNA 2008 Sequence of accounts; see table 2) is the sum of all revenues from taxes on production received by the government, plus the property income received by the government, minus the property income paid by the government. Two remarks are in order. First, because the government sector has some market activity, the primary income of the government sector also includes a small “net operating surplus” component. In order to simplify exposition and tables, we choose to treat the small net operating surplus of the government as “taxes on production”; see formulas in table 2. Second, by convention, the rental value of real assets owned and used by the government does not generate net primary income. The real assets owned and used by the government are assumed

to have a 0% net-of-depreciation return, and a gross-of-depreciation return equal to the rate of capital depreciation (this convention could and probably should be changed in the future; but at this stage we take it as given). By contrast, the real assets owned by the government but rented to other sectors generate net operating surplus. We treat the flow of operating surplus generated by these assets as production taxes (see above remark).

Finally, primary income of the non-profit sector (i.e. non-profit institutions serving households (NPISH), as defined in SNA categories, which make less than 1% of national income according to the example chosen in the SNA 2008 Sequence of accounts; see table 2) is the sum of the net operating surplus of the non-profit sector, plus property income received by the non-profit sector, minus the property income paid by the non-profit sector. The net operating surplus of the non-profit sector is equal to the rental value of the real assets rented by non-profits to other sectors. Just like for the government, the rental value of the assets owned and used by nonprofits does not generate net primary income (these assets are assumed to have a 0% net-of-depreciation return, and a gross-of-depreciation return equal to the rate of capital depreciation).

Section 3.4. Pre-tax factor income

Pre-tax factor income, which for simplicity we sometime refer to as “factor income”, is equal to the sum of all pre-tax income flows accruing directly or indirectly to the owners of the production factors, labor and capital, before taking into account the operation of

the tax/transfer system (including indirect taxes), and before taking into account the operation of the pension system.

The relation between pre-tax factor income and national income is presented on table 3. By construction, aggregate pre-tax factor income is exactly equal to aggregate national income, and can be broken down into personal factor income, government factor income, and non-profit factor income. Government and non-profit factor income are defined as the difference between the property income received by the government and non-profit sectors and the property income paid by the government and non-profit sectors. In practice, government interest payments often exceed government property income receipts in most of today's developed economies, so that government factor income is often negative and personal factor income tends to exceed national income. E.g. personal factor income is equal to 101% of national income according to the example chosen in the SNA 2008 Sequence of accounts (see table 3). Conversely, in countries where the government receives substantial positive property income (via a large public sector or sovereign wealth fund, for instance in China or Norway), government factor income tends to be substantially positive, personal factor income could be much less than national income. In section 5 below, we explain how we attribute government and non-profit factor income to individuals.

As one can see from table 3, personal factor income can also be computed as the sum of primary income of the household sector, the primary income of the corporate sector

(i.e. undistributed profits),[18] and the revenues from taxes on production received by the government. Three remarks must be made here.

First, the key reason for adding undistributed profits (or at least a fraction of them) to personal factor income is because undistributed profits should be considered as income for the owners of corporations. Undistributed profits are an income flow in the Hicksian sense: they make the owners of corporations wealthier. Depending on the tax system, individual shareholders may prefer to accumulate profits in corporations rather than to receive dividends (e.g., because this may allow them to realize capital gains by selling shares at a later stage, and by doing so they might pay less taxes than what they would have paid on the corresponding dividends). The best way to correct for this is and to make our estimates comparable over time and across countries to add undistributed profits to personal factor income, at least in part.

The question is whether it is justified to add 100% of undistributed profits to personal factor income. To the extent that the government owns a negligible part of the corporate sector, and to the extent that the foreign asset position of the country is broadly balanced (the fraction of domestic corporations owned by the rest of the world is often close to what domestic households own in corporations in the rest of the world, and undistributed profits represent a similar share of total profits in domestic and foreign corporations), then such an imputation strategy might be justified, at least as a first approximation. However in countries where the government owns a significant fraction of the domestic corporate sector (and/or where the rest of the world owns a

[18] Here “undistributed profits” refer to the net primary income of corporations, i.e. pre-tax undistributed profits. In practice the net primary income of corporations is equal to the sum of retained earnings, corporate income tax paid and current transfers.

fraction of the domestic corporate sector that is significantly higher or smaller than what domestic sectors own in the rest of the world), it is preferable to use more balanced imputation strategies. We discuss these issues in the country specific chapters (see in particular Piketty-Yang-Zucman 2017 for the case of China, where the government owns a very large part of the domestic corporate sector, so that it would make little sense to attribute 100% of undistributed profits to personal factor income). One should also stress that official national accounts guidelines are not fully consistent in the way they treat undistributed profits in foreign-owned corporations.[19]

Next, the key reason for adding production taxes to personal factor income is because the frontier between production taxes (D2) and direct income and wealth taxes (D5) is somewhat arbitrary, i.e. it is unclear why we should deduct the former and not the later. For the purpose of making comparisons over time and across countries, it makes more sense to look at the distribution of income before the deduction of any tax, either production taxes of other taxes. Of course one needs to make assumptions regarding tax incidence in order to impute production taxes (just like for other taxes). On table 3, we make the simplest possible benchmark assumption: we assume that production taxes fall proportionally on the different income categories. We will discuss more sophisticated, alternative assumptions in section 5 below.

[19] In particular, undistributed profits on direct investment abroad (defined as more than 10% ownership) are always added to net primary income (Reinvested earnings on foreign direct investment received, D43 in the SNA 2008 classification); conversely, retained earnings of domestic corporations which are owned more than 10% by foreigners are subtracted from net primary income (D43 paid). The only problem is for retained earnings on portfolio investment (less than 10% ownership), which are not added/subtracted to/from primary income. Until the early 1980s, more than 90% of U.S. foreign equity assets are FDI investments (and less than 10% are portfolio assets), so that almost all foreign retained earnings are included in national income. But since the mid 1980s the share of FDI gradually dropped and is currently close to 50%, so that national income misses about 50% of foreign retained earnings. This can be viewed as a shortcoming of national accounts guidelines, which did not foresee the rise of porfolio foreign investment since the 1990s-2000s.

We also report a decomposition of factor income into labor income and capital income on table 3, assuming a simple 70-30 split of self-employment income into a labor share and capital share (we discuss alternative assumptions in section 5 below). With the examples chosen in the SNA 2008 Sequence of accounts, we come to a 82%-18% split of pre-tax factor income into labor and capital (see table 3).

One positive aspect of the factor income concept is that it is relatively easy to compute using national accounts data, and it is reasonably homogenous across countries. The main drawback, however, is that old-age individuals generally have little factor income, so that cross-sectional inequality of factor income looks artificially large in countries and time periods with large old-age population. This is illustrated by figure 1, where we report the age profile of factor income using prototype DINA files for France 2006 (see Landais, Piketty and Saez, 2011).

Our series correct for this in two ways: first, by providing separate factor income inequality series within each age group (in particular, within the working-age population); next, and most importantly, by computing inequality series using another income concept, namely pre-tax national income.

Section 3.5. Pre-tax national income

Pre-tax national income, which for simplicity we sometime refer to as "pre-tax income", is equal to the sum of all pre-tax income flows accruing to the individual owners of the

production factors, labor and capital, before taking into account the operation of the tax/transfer system, but after taking into account the operation of the pension system.

The relation between pre-tax national income and pre-tax factor income is presented on table 4. By construction, they are both equal to national income at the aggregate level. But they are not the same at the individual level and in terms of distribution. The central difference between pre-tax factor income and pre-tax national income is the treatment of pensions, which are counted on a contribution basis by pre-tax factor income and on a distribution basis by pre-tax national income. We tend to favor the "pre-tax national income" concept, and we view our "pre-tax national income" inequality series as our benchmark series for pre-tax inequality. We stress however that the "pre-tax factor income" inequality series also provide useful and complementary information. To the extent possible, both series should be estimated and computed for all countries.

The key reason why we tend to prefer the "pre-tax national income" series is that they are less strongly affected than pre-tax factor income inequality by the age structure of the population. As was mentioned above, pre-tax factor income inequality is artificially large in economies with a large retired population (even if the pension system provides full replacement). In contrast, we aim to define pre-tax national income so as to satisfy the following neutrality condition: in a hypothetical economy with 100% replacement rates for pensioners (whether this comes from a compulsory pay-as-you-go pension system, or a voluntary funded system, or any combination between the two), the cross-sectional inequality of pre-tax national income should be the same whether it is measured within the entire population (including pensioners) or within the working-age

population. In particular, if there is no labor income inequality whatsoever between workers (think of a representative agent OLG economy with equal wages), then with a pension system with full replacement there should be no cross-sectional inequality of pretax income within the entire population.

The way one can compute pre-tax national income using *SNA 2008* classifications is described on table 4. Several difficulties should be pointed out. In particular, there is an issue as to whether we should apply the distribution principle to all social insurance benefits and contributions (“broad” definition), or only to the pension component (“pension-based” definition). On table 4 we favor the “broad” definition, but this is debatable.

In the "broad" definition, we deduct all social contributions (and not only the pension contributions), i.e., D61 (as recorded by *SNA 2008*), and we add all the social insurance benefits (and not only the pension benefits), i.e., the sum of D621 and D622. In practice, pensions generally represent the vast majority of social contributions and social insurance benefits, and the main non-pension social insurance benefits are unemployment insurance benefits (which in many countries is treated as "replacement income", together with pensions, and are subject to the income tax as regular income, at least when they exceed a certain level).

In the "pension-based" definition, one should instead deduct from factor income the pension contributions (as defined by *SNA 2008*, i.e., the sum of D6111, D6121, D6131, D6141, so as to include contributions made by employers and households to public and private pension systems) and the investment income payable to pension

entitlements (D442), and add the pension benefits – i.e. the sum of D6211 (social security pension benefits, i.e., public pensions) and D6221 (other social insurance pension benefits, i.e. private and occupational pensions).

Note that *SNA 2008* distinguishes between social insurance benefits (i.e., the sum of D621, social security benefits in cash, and D622, other social benefits) and social assistance benefits in cash (D623). The difference is that entitlements to social insurance benefits are based upon contributions, while entitlements to social assistance benefits are not. We exclude social assistance benefits in cash (D623), as well as social transfers in kind (D63), from pretax income (broad definition), and include them solely in post-tax income (see section 3.4 below).

Should non-pension social insurance benefits be included in pretax income? In our view, there are costs and benefits associated to both the "pension based" and the "broad" definitions of pretax income. We generally recommend using the "broad" definition, primarily because it is less data intensive and easier to implement on an international basis. In most countries, national accounts are currently not available with the detailed classifications defined by *SNA 2008*. Typically, the decomposition between pension and non-pension social insurance benefits and contributions is generally not available.[20] When it is available, or when other data sources allow doing the decomposition, we recommend to apply both definitions of pretax income and compare the level and trends in inequality.

[20] Note that "investment income payable to pension entitlements" (D442) is often not available in existing national accounts (one typically observes "investment income disbursements", i.e. D44), and therefore needs to be estimated using other sources.

An additional reason for using the "broad" definition is that one might want to neutralize the impact of "unemployment risk" on inequality, in the same manner as we neutralize the impact of "old age risk". In a number of countries, unemployment insurance benefits are approximately proportional to contributions, in the same way as pensions, in which case it makes sense to treat them together with pensions. More generally, the "broad" definition aims to include all forms of "social insurance income" (or "replacement income") into pre-tax income.

We represent on figure 2 the age profile of pre-tax national income (broad definition) observed in France (see Garbinti, Goupille and Piketty, 2017). Unsurprisingly, this looks more balanced than the age profile of factor income (see figure 1). As a consequence, inequality measures that are based upon pre-tax national income are much less affected by changes in the age structure of the population than those using pre-tax factor income.

One difficulty with the “broad” definition is that in practice the frontier between contributions-based social insurance benefits and non-contributions-based social assistance benefits is not entirely clear: some benefits classified as social insurance benefits by *SNA 2008* and/or by national accounts statisticians clearly have a strong redistributive component, in which case it might be justified to make corrections and to estimate several variants. For instance, in the prototype DINA for France estimated by Landais-Piketty-Saez (2011), the choice was made to exclude family benefits from "social security benefits in cash" (D621) and treat them as part of "social assistance benefits in cash" (D623), on the basis that family benefits bear little relation with contributions.

Another difficulty has to do with the possible imbalance between contributions and benefits. In the example provided in the SNA 2008 Sequence of Accounts, contributions and benefits are almost equal, but there is a small surplus of contributions over benefits (see table 4). In some countries one might observe major imbalance between contributions and benefits, partly for "good" economic reasons (e.g., it could be that the pension system is temporarily accumulating surpluses or deficits, due to demographic changes), and partly for "bad" accounting reasons (e.g., it could be that the pension system is partly financed by general tax revenues rather than by social contributions, or conversely that social contributions – as recorded by national accounts – finance public spending that are not counted as social insurance benefits). In both cases, we include this surplus (positive or negative) in our definition of pre-tax national income, so that aggregate pre-tax national income is exactly equal to aggregate national income (see table 4). We discuss in section 5 how to attribute this surplus to individuals.

It is worth noting that even if the pension system is in steady-state, and even in the absence of any "bad" accounting reason, there could exist some structural gap between contributions (and investment income) on the one hand, and pension distributions on the other hand. Assume the economy is in steady-state growth (fixed demographic and productivity growth rates, with a stable age structure) with a total growth rate equal to g=n+h (the sum of demographic and productivity growth), and an average return to capital equal to r. With a pay-as-you-go pension system, then contributions are by definition equal to pensions, so that there is no surplus. However with a funded pension system with total steady-state pension wealth equal to $W_{Pt}=\beta_P Y_t$

(where Y_t is national income, growing at rate g; W_{Pt} is pension wealth, also growing at rate g; and β_P is the steady-state pension wealth-national income ratio), one can immediately see that contributions (and accrued investment income) exceed pension distributions by gW_{Pt}. So for instance if g=2% and steady-state pension wealth represents 200% of factor income, then in steady-state the surplus of the pension system will be 4% of national income in a country with funded pensions (and 0% in a country with pay-as-you-go pensions). For instance, in the United States, we find a surplus of 5-10% of national income in recent decades (this ratio is abnormally high because of reserve accumulation and should fall below 5% as we approach steady-state; see Piketty, Saez and Zucman, 2016), while it is close to 0% in France (see Garbinti, Goupille and Piketty, 2016).

In practice, there are all sorts of out-of-steady-state reasons why the pension system is not exactly balanced, and we feel that in order to make comparisons over time and across countries it is preferable to include the surplus of the pension and social insurance system into pre-tax national income. Note however that in the case of a steady-state surplus due to difference in pension systems (funded vs pay-as-you-go) it is not entirely clear whether one should do this.[21]

[21] E.g. assume an open economy with a fixed world rate of return r. If we take everything else as given (in particular if we take other saving motives as given), then the pension-fund country will accumulate more wealth and will therefore have a national income that exceeds that of the pay-as-you-go country by rW_{Pt}. However the pension-fund country needs to save an extra amount equal to gW_{Pt} in order to sustain this higher wealth accumulation, i.e. each year the sum of pension contributions and pension investment income needs to exceed pension distributions by gW_{Pt} (while in the pay-as-you-go country there is no such steady-state pension surplus: contributions are equal to distributions). In other terms, national income is higher in the pension-fund country by rW_{Pt}, but if we deduct these extra savings the real difference in terms of steady-state resources available for consumption and investment is $(r-g)W_{Pt}$. As is well-known, funded pensions make sense only when the dynamic efficiency condition r>g is satisfied (otherwise we are in a situation of excessive capital accumulation).

Section 3.6. Post-tax national income and post-tax disposable income

We define two notions of post-tax post-transfer income, namely post-tax national income and post-tax disposable income (see table 5).

Post-tax disposable income is defined as pre-tax national income, minus all taxes on production, income and wealth, plus social assistance benefits in cash (D623).

In order to compute post-tax national income, we add social transfers in kind (D63), which according to *SNA 2008* includes in-kind transfers such as education and health expenditures (and more generally all transfers of goods and services by the government and non-profit sectors which can be consumed at the individual level), and we also include "collective consumption expenditure" (P32), which according to *SNA 2008* includes public spending such as national defense and street lighting (and more generally all provision of goods and services by the government and non-profit sectors which can be consumed only at the collective level). Needless to say, attributing such items to individuals is bound to be approximate and exploratory (see the discussion in section 5 below), which is why we also report results for post-tax disposable income. Finally we also include government primary surplus (positive or negative) into post-tax national income, so that aggregate post-tax national income again coincides with aggregate national income (see the discussion in section 5 on how we allocate the government surplus to individuals). Note that this notion of post-tax national income corresponds approximately to what is defined by *SNA 2008* as adjusted national disposable income (except that we do not deduct current international transfers such as remittances).

Section 4. Wealth concepts

We now define the various concepts of wealth, assets and rates of return that we use in WID.world. In the same way as for the income concepts, our wealth concepts are defined using the latest international guidelines regarding national accounts (*SNA 2008*). We again recommend that readers look at the *DINA concepts* excel file where we provide formulas relating our definitions to the "Sequence of accounts" excel file and to the SNA 2008 classification codes. We begin with the definition of personal and private wealth. We then move to rates of return by class of assets (these definitions will play a major role in order to apply the income capitalization method to income tax data; see the discussion below). Finally we present the definitions of national, public and foreign wealth that we use in WID.world.

Section 4.1. Personal wealth and private wealth

We define personal wealth as the net wealth of the household sector, i.e. the sum of non-financial and financial assets owned by households, minus their financial liabilities, as defined by *SNA 2008*. The details of the computations are given on table 6, where we also provide a number of decomposition into different classes of assets.

Our basic decomposition includes four classes of assets and liabilities: housing assets; business assets (and other non-financial assets); financial assets; and liabilities. Housing assets are defined as the sum of the market value of dwellings and land underlying dwellings (in practice, it is generally easier to measure the sum - e.g., in observed real estate transactions - than the two components separately). Business

assets (and other non-financial assets) are simply defined as the difference between total non-financial assets and housing assets.

Note that existing national balance sheets do not always provide separate estimates for the different uses of land. In the basic classification codes used in *SNA 2008,* land appears as a single asset (classification code AN.211). In the detailed *ESA 2010* classification codes, land (AN.211) is broken down into "land underlying buildings and structures" (AN.2111), "land under cultivation" (AN.2112), “recreational land and associated water surfaces” (AN.2113), and “other land and associated water surfaces” (AN.2114). Many national statistical agencies also break down "land underlying buildings and structures" (AN.2111) into "land underlying dwellings" (AN.21111) and "other land underlying buildings and structures" (AN.21119). When this latter decomposition is not available, we recommend splitting the land value in proportion to value of dwellings and other buildings and structures (see the *DINA concepts* excel file where we provide formulas for the example of France).

Whenever possible, we also recommend to break down business assets (and other non-financial assets) into agricultural land (AN.2112) and other domestic capital (i.e. all non-financial assets except housing and agricultural land) (see table 6). Although agricultural land is now a negligible part of assets in the balance sheet developed countries, it obviously played a very large historical role, and still plays an important role in developing countries with a large agricultural sector, so it is important to provide this decomposition. More generally, the study of the comparative structure of land value and of the long-run decomposition between rural and urban land is a critical and

complex issue that would deserve further attention, and which the WID could contribute to clarify.[22]

Finally, we aim to split financial assets into three categories (see table 6): currency, deposits, bonds and loans (the sum of AF.1, AF.2, AF.3, AF.4, AF.7 and AF.8); equity and investment fund shares (AF.5); life insurance and pension funds (AF.6). For some countries, it might be possible and justified to use more detailed breakdowns. We return to this below when we discuss the computation of rates of return and the implementation of the income capitalization method.

Note that in some countries, available balance sheets include the assets and liabilities of the non-profit sector together with those of the household sector. In such cases, we cannot compute personal wealth (as defined on table 6), and we can only compute private wealth (as defined by the sum of personal wealth and non-profit wealth). Given that non-profit wealth can represent a non-negligible fraction of private wealth, we recommend to estimate at least some approximate break-down of private wealth into personal and non-profit wealth, for instance by using the decomposition of capital income flows (which are usually available separately for the household and non-profit sectors). When balance sheets are available separately for the household sector and the non-profit sector, then one can easily provide for non-profit wealth the same

[22] Needless to say, the frontier between the pure land value and the value of the capital accumulated on the land (or the value of improvements made to the land) is often difficult to estimate. According to *SNA 2008* and *ESA 2010*, whenever it is impossible to separate land and building value, all value is allocated to the biggest part. Also, note that that "other buildings and structures" (AN.112) are broken down into "buildings other than dwellings" (AN.1121), "other structures" (AN.1122), and "land improvement" (AN.1123). However AN.1123 - when available - is typically very small, probably because it only takes into account the recent land improvement, not the entire historical sequence of non-human and human investment and improvement that made rural and urban land valuable since the beginning of mankind. For further discussion, see Piketty-Zucman (2014) and Piketty (2014, chap.6).

decomposition as for personal wealth (see table 10), and we can compute private wealth as the sum of two (see table 11).

Section 4.2. Rates of return

We provide on table 7, table 8 and table 9 computations of average rates of return by asset classes using *SNA 2008* classification codes and the concepts of income and wealth that we defined in the previous tables. These average rates of return are computed by linking classifications of assets and asset income flows, and by dividing the latter by the former. This will play an important role when we discuss the income capitalization method, which can be used to estimate the distribution of wealth from the distribution of capital income flows (see section 5 below).

On table 7 we start with the classification with four assets/liabilities: housing assets, business assets (and other non-financial assets), financial assets, and liabilities. Using the SNA 2008 "Sequences of accounts" tables as an example, we find average rates of return that are relatively close for the four assets/liabilities (between 5.6% and 7.7%). Note that these are pretax rates of returns, and that undistributed profits (including corporate income tax payments) were attributed to financial assets. Note also that all production taxes were attributed to factor labor and capital income flows in proportion to each income flow (which might be acceptable as a first approximation, but which could be improved; e.g., property taxes could be attributed to housing; see the discussion in section 5 below).[23] Note that on WID.world we provide separate series for production taxes and factor-price national income (ie. we do not impute production

[23] To be added: discussion on imputed interest income.

taxes to labor and capital income), so that users can make their own imputations if they want to.

On table 8 we break down the average rate of return on financial assets into the three financial asset categories that were defined above. We make three different assumptions regarding undistributed profits: we attribute them either to equity (and investment fund shares), to the sum of equity (and investment fund shares) and life insurance and pension funds, or the sum of all financial assets. On table 9 we provide the same computations using 2013 national accounts for France rather than the SNA 2008 "Sequence of accounts" table. In countries, such as the United States, where there exists information on the composition of the wealth of pension funds and life insurance companies (i.e., what fraction is invested in equities vs. other assets), then the best solution might be to allocate undistributed profits to equity and the fraction of pension funds & life insurance companies' wealth which is invested in equities. Otherwise, it may make sense to attribute undistributed profits to all financial assets, but this is an issue on which we feel we need to perform more sensitivity and robustness checks for more countries. We will return to these issues in section 5 when we discuss the question of corporate tax incidence.

Finally, note that the change of national accounts system from *SNA 1993* to *SNA 2008* involved a number of generally minor changes, but which in some cases might have significant consequences for the definitions of the different asset-level rates of return. The classifications of financial assets were virtually unchanged in the new system, so that we can define our three main categories of financial assets in the same manner in the two systems: deposits, currency, bonds and loans (sum of AF1, AF2, AF3, AF4,

AF7, AF8); equity and mutual funds (AF5); life insurance and pension funds (AF6).[24] However the classifications of property income flows were changed in a significant way: namely, D44 "Investment income disbursements" now includes D443 "Investment income attributable to collective investment funds share holders", in spite of the fact that the corresponding assets are still included with equities (AF5). This flow of property income going to mutual funds and other investment funds (other than life insurance and pension funds) used to be included into D42 (together with dividends and other property income flows going to AF5-type financial assets). When the detailed series are available, we recommend to reattribute D443 to the flow of property income going to equity and mutual funds (see the discussion of Garbinti, Goupille and Piketty (2016) in the case of France). In some cases, one may also prefer to isolate "deposits and currency" (AF2) within the broader asset category "deposits, currency, bonds and loans" (see also Garbinti, Goupille and Piketty (2016)). In all cases, we recommend to perform multiple sensitivity tests when applying the income capitalization method (see discussion in section 5 below and in the country-specific papers).

Section 4.3. Public wealth

In order to define public wealth, residual corporate wealth, and national wealth (for which we provide what we view as two complementary definitions: market-value and book-value), it is useful to introduce some formal notations.

[24] The main change in SNA 2008 is the introduction of AF7 "Financial derivatives" (former AF7 "Other accounts" becomes AF8).

Private wealth W_{pt} is the net wealth (assets minus liabilities) of households and non-profit institutions serving households and can be broken down as follows:

$$W_{pt} = K_{pt} + F_{pt} - L_{pt}$$

With: K_{pt} = non-financial assets owned by private sector (which can be further decomposed into $K_{pt} = H_{pt} + A_{pt} + D_{pt}$, i.e. housing assets + agricultural land + other domestic capital)[25]

F_{pt} = financial assets owned by private sector (which can be further decomposed into $F_{pt} = C_{pt} + E_{pt} + I_{pt}$, i.e. currency-deposits-bonds-loans + equities-shares-offshore + pension-funds-life-insurance)

L_{pt} = financial liabilities (debt, bonds, loans etc.) of the private sector

In the same manner, we define public wealth as the net wealth of the government sector, i.e., the sum of non-financial and financial assets owned by government entities, minus their financial liabilities, as defined by *SNA 2008*. The details of the computations are given on table 12, where we also provide the same decomposition into different classes of assets as for private wealth. I.e. government (or public) wealth W_{gt} can be decomposed as follows:

$$W_{gt} = K_{gt} + F_{gt} - L_{gt}$$

[25] Given the limitations in data availability, we typically include in other domestic capital all forms of non-financial assets recorded in SNA guidelines other than housing assets and agricultural land. This includes in particular natural resources (other than agricultural land and land underlying dwellings, which is included in housing values) and intellectual property (which is included since SNA 2008). However, whenever the full decomposition is available, we aim to separate other domestic capital from natural ressources and intellectual property.

With: $K_{gt} = H_{gt} + A_{gt} + D_{gt}$ = non-financial assets owned by government sector

$F_{gt} = C_{gt} + E_{gt} + I_{gt}$ = financial assets owned by government sector

L_{gt} = financial liabilities (debt, bonds, loans, etc.) of the government sector

Section 4.4. Residual corporate wealth

We do the same on table 13 for the corporate sector. As is well-known, there are two ways to compute the value of corporations. One can use the market equity value of corporations EV_{ct}, and the book-value of corporations BV_t , which is defined as the difference between assets and (non-equity) liabilities:

$$BV_t = K_{ct} + F_{ct} - DL_{ct}$$

With: $K_{ct} = H_{ct} + A_{ct} + D_{ct}$ = non-financial assets owned by corporate sector

$F_{ct} = C_{ct} + E_{ct} + I_{ct}$ = financial assets owned by corporate sector

DL_{ct} = debt liabilities of corporate sector (i.e. non-equity corporate liabilities: debt, bonds, loans, etc.) of corporate sector

EV_{ct} = market equity value of corporate sector (equity corporate liabilities, i.e. total market equity value of domestic quoted and unquoted corporations)[26]

$L_{ct} = DL_{ct} + EV_{ct}$ = total financial liabilities of the corporate sector (debt and equity)

One can define the Tobin's Q ratio as the ratio Q_t between the market equity value of corporations EV_{ct} and the book value of corporations BV_t , and residual corporate wealth W_{ct} as the difference between the book value and the market value:

[26] Unquoted shares are typically valued on the basis of observed market prices for comparable, publicly traded companies.

$$Q_t = EV_{ct} / BV_t$$

$$W_{ct} = BV_t - EV_{ct} = K_{ct} + F_{ct} - L_{ct}$$

Residual corporate wealth W_{ct} should be viewed as "residual" in the following sense. In practice, the corporate sector is owned in part by the other two domestic sectors (private sector or government sector) and in part by the rest of the world (foreign sector), so the value of corporations – as measured by their market equity value – is already included in the financial assets and therefore the net wealth of these other sectors.

In case Q_t is equal to one, i.e. if market value and book value are the same, then by construction residual corporate wealth is equal to zero: the full value of corporations is already included in private and public wealth so there is nothing to add.

In case Q_t is less than one, which is often the case in practice (e.g. in Germany, Japan or France, as well as in the UK and the US until the 1990s-2000s), then residual corporate wealth is positive: corporations own assets that are undervalued on the stock market (as compared to their book value), possibly because of various measurement errors (either in book values, market values, or both), or because shareholders have to share power with other stakeholders and cannot easily liquidate all company assets (even if they wanted to).

Conversely, in case Q_t is higher than one, which happens for certain periods and countries (e.g. in the UK and the US since the 1990s-2000s, at least prior to the 2008

crisis), then residual corporate wealth is negative: corporations enjoy stock market values that exceed the value of the assets recorded on their books, possibly because of various measurement errors, or because the market perceives that they benefit from unrecorded immaterial assets, rights, market power or reputation that are likely to boost their profitability.[27]

Section 4.5. Market-value vs book-value national wealth

We provide two complementary definitions of national wealth: market-value and book-value (see table 14 for the corresponding decompositions).

First, we define market-value national wealth W_{nt} as the sum of private wealth and government wealth (i.e. we ignore residual corporate wealth, which cannot be directly attributed either to private individuals or to the government):[28]

$$W_{nt} = W_{pt} + W_{gt} = K_{nt} + NFA_{nt}$$

With: $NFA_{nt} = F_{nt} - L_{nt}$ = net foreign assets owned by domestic sectors (with $F_{nt} = F_{pt} + F_{gt} + F_{ct}$ = total financial assets owned by private, government and corporate sectors in the national economy and in the rest of the world, and $L_{nt} = L_{pt} + L_{gt} + L_{ct}$ = total financial liabilities of private, government and corporate sectors); this can also be

[27] See Piketty and Zucman (2014, Figure 92) for the evolution of Q ratios over the 1970-2010 period.
[28] Conceptually, the issue as to whether residual corporate wealth should be attributed to private individuals or to the government is related to the issue of attribution of non-profit wealth (which we attribute to the private sector, largely because it is actually not distinguished from household wealth in many countries). The main rationale for looking at market-value national wealth is the possibility of measurement error for non-financial corporate assets and the view that stock market values might provide a more accurate evaluation of the "real" value of corporations (which is far from clear).

decomposed as $NFA_{nt} = GFA_{nt} - GFL_{nt}$, with GFA_{nt} = gross financial foreign assets (i.e. gross financial assets owned by domestic sectors in the rest of the world) and GFL_{nt} = gross financial foreign liabilities (i.e. gross financial assets owned by the rest of the world in domestic sectors, which can itself be decomposed between equity and non-equity foreign liabilities GFE_{nt} and GFN_{nt})

K_{nt} = market-value domestic capital = sum of domestic non-financial assets owned by private, government and corporate sectors, i.e. $K_{nt} = K_{pt} + K_{gt} + K_{ct}^*$, with $K_{ct}^* = K_{ct} - W_{ct}$ = corrected value of corporate non-financial assets, using the implicit stock-market valuation for non-financial corporate assets and assuming stock market prices provide the most accurate valuation for these assets. We can also decompose K_{ct}^* into $K_{ct}^* = H_{ct}^* + A_{ct}^* + D_{ct}^*$ by assuming the same proportional correction for all non-financial corporate assets, i.e. $H_{ct}^* = H_{ct} \ K_{ct}^*/K_{ct}$; $A_{ct}^* = A_{ct} \ K_{ct}^*/K_{ct}$; $D_{ct}^* = D_{ct} \ K_{ct}^*/K_{ct}$. We can then decompose market-value domestic capital K_{nt}^* into its various subcomponents: $K_{nt}^* = H_{nt}^* + A_{nt}^* + D_{nt}^*$, with $H_{nt}^* = H_{pt} + H_{gt} + H_{ct}^*$, $A_{nt}^* = A_{pt} + A_{gt} + A_{ct}^*$, $D_{nt}^* = D_{pt} + D_{gt} + D_{ct}^*$.

Note that by convention, the rest of the world does not directly own non-financial assets in the national economy. In case foreign residents – either private individuals, governments or corporations – own domestic non-financial assets, this is accounted for as if they own financial assets in a domestic fictitious corporation, which then owns the domestic non-financial assets.[29]

[29] The same rule applies to foreign non-financial assets owned by domestic sectors: they appear entirely as foreign financial assets. By construction all non-financial assets owned by domestic sectors are domestic non-financial assets.

Next, we define book-value national wealth W_{bt} as the sum of private wealth, government wealth and residual corporate wealth:

$$W_{bt} = W_{pt} + W_{gt} + W_{ct} = K_{bt} + NFA_{nt}$$

With: K_{bt} = book-value domestic capital = total domestic non-financial assets = sum of domestic non-financial assets owned by private, government and corporate sectors, i.e. $K_{bt} = K_{pt} + K_{gt} + K_{ct}$, which can also be decomposed as $K_{bt} = H_{nt} + A_{nt} + D_{nt}$ = housing + agricultural land + other domestic capital (with $H_{nt} = H_{pt} + H_{gt} + H_{ct}$, $A_{nt} = A_{pt} + A_{gt} + A_{ct}$, $D_{nt} = D_{pt} + D_{gt} + D_{ct}$)

An index of domestic financial intermediation can be defined as follows:

$$DFI_t = L_{nt}/K_{nt} = (F_{nt} - GFA_{nt} + GFL_{nt})/K_{nt}$$

This index measures the quantities of domestic financial assets and liabilities that are generated to organize the ownership of a given unit of domestic real capital.

An index of financial foreign ownership can be defined as follows:

$$FFO_t = GFL_{nt}/L_{nt} = GFL_{nt}/(F_{nt} - GFA_{nt} + GFL_{nt})$$

This index measures the fraction of domestic financial assets and liabilities that are owned by the rest of the world.

An alternative index is the fraction of equity foreign ownership in total domestic equity:

$$EFO_t = GFE_{nt}/EV_{nt}$$

One can also define an index of real gross foreign ownership as follows:

$$GFO_t = GFL_{nt}/K_{nt} = FFO_t \times DFI_t$$

It should be noted however that in economies with very high domestic financialisation (high index of domestic financial intermediation DFI_t), the index GFOt can be very high. In particular it can be higher than 100% in countries with large international financial centers such as Britain.

Finally one can define an index of real net foreign ownership as follows:

$$RFO_t = -NFA_{nt}/K_{nt}$$

I.e. RFO_t measures the equivalent fraction of the domestic capital stock owned by the rest of the world in net terms (that is, given what residents own in the rest of the world).

We stress again that whether residual corporate wealth should be included in national wealth is really a matter of perspective (see Piketty and Zucman (2014) for a more detailed discussion). Excluding residual corporate wealth", as in our benchmark measure of "market-value national wealth", means that we value corporate assets at market value, as reflected in the prices of corporate bonds and corporate equities. This

can be justified by the view that market values of corporations are better estimates than book values of corporations, for instance because different forms of non-financial assets, in particular coming from intangible investment, are not well taken into account in existing balance sheets. If systematic deviations of Tobin's Q from unity only reflect measurement errors, then they should be ignored, and our benchmark “market-value national wealth” definition is the most appropriate.

In our alternative definition of national wealth, “book-value national wealth”, corporations are not valued using at market prices, but are valued according to what their assets are recorded to be worth in the corporate sector’s balance sheet. “Book-value national wealth” is equal to the sum of all the non-financial assets of all domestic sectors, plus the net foreign asset position. This definition can be meaningful if deviations of Tobin’s Q from unity do not reflect measurement errors only, but also real changes in the balance of power between the various stakeholders of corporations. Tobin's Q lower than one (positive net corporate wealth) might reflect the fact that stakeholders other than shareholders partly control companies’ income flows (like in Germany). Conversely, Tobin's Q higher than one (negative net corporate wealth) might reflect the fact that shareholders are able to extract high rent from companies, maybe because the legal system is very favorable to them (like in the U.S. and in the U.K.). In these cases, “book-value national wealth” can be interpreted as capturing the value corporations from the viewpoint not of firms’ owners (as reflected in equity and bond prices, and captured by “national wealth”), but of all stakeholders of the firms.

Our view is that both approaches to national wealth are useful and complementary, and that collecting more data series from more countries using both approaches might help us to better understand their respective relevance and limitations.

We refer to Piketty and Zucman (2014, figure 3) and Piketty (2014, figure 3.1 and figure 3.2) for long-run decompositions of national wealth as the sum of agricultural land, housing assets, other domestic capital assets, and net foreign wealth.

Finally, note that although we are primarily interested in estimating the distribution of personal wealth between private individuals (see sections 5-6 below), it could also be interesting in some cases to construct estimates of the distribution of national wealth between individuals. E.g. in a country with large public wealth (such as Norway), it may make sense to attribute public wealth to private individuals (otherwise the residents of Norway might artificially appear to hold very little wealth as compared to other countries). The same issue arises for countries with significant negative public wealth (large public debt relative to public assets). One way to attribute public wealth (positive or negative) to private individuals would be in proportion to tax liabilities - or in proportion to spending entitlements (e.g. rights to pensions financed out of public sovereign funds), which might be substantially different. In countries where sovereign wealth is controlled by a smaller group of the population (i.e. in Gulf countries), one might choose to use specific imputation methods. These are important and complex issues for the future.

Section 5. Basic imputations Methods using Income Tax Micro-Files

We now move to the presentation of our basic imputation methods. In this section, we implicitly assume that we have access to high-quality income tax micro files including reliable annual information on individual flows of both labor and capital incomes and covering the entire population, together with high-quality income and wealth surveys, so that these two data sources can be combined with national accounts to compute homogenous estimates of the distribution of the income and wealth using concepts consistent with national accounts.

While this assumption is satisfied in a number of developed countries today and in recent decades (e.g. in the U.S. we have access to high-quality fiscal micro-files covering almost the entire population since 1962, which we can use to construct our DINA series; see Piketty, Saez and Zucman, 2016 for detailed estimates and presentation of the imputation methods; in France we have access to similar micro-files since 1970; see Garbinti, Goupille and Piketty, 2016, 2017). Unfortunately, this is not the case in most emerging countries today (such as China) or for developed countries in earlier time periods (e.g. in the U.S. before 1962 or in France before 1970, we do not have access to income tax micro-files). In sections 7-8, we will discuss the methods that can be used in the case of countries and time periods with more limited data sources, typically with income tax tabulations instead of micro-files, and/or with income tax data covering only a subset of the population rather than the entire population (this would apply to the case of China today, and to some extent to the U.S. and France prior to 1962 or 1970). We will also discuss in section 7 how initial WTID

series using a fiscal income concept can be corrected so as to be more directly comparable to new DINA series.

When high-quality income tax micro files are available, the basic imputation and estimation methods that we use to produce DINA series are relatively straightforward. That is, we start from these income tax micro files, we scale up fiscal income flows up to national-accounts-based income concepts (using the concepts of pre-tax factor income, pre-tax national income, post-tax disposable income and post-tax national income and disposable income defined in section 3), and we use the income capitalization method in order to recover the wealth distribution from capital income flows (using the concepts of wealth, assets and rates of return defined in section 4), in conjunction with income and wealth surveys and other data sources on wealth (in particular inheritance tax returns, when available, as well as wealth rankings) in order to cover assets that do not generate taxable capital income flows and to ensure the robustness of the income capitalization method. We describe the main steps below, and we discuss in section 6 how other data sources on wealth (including inheritance data) should be reconciled and combined with the income capitalization method. More details are provided in the country-specific papers (see in particular Piketty-Saez-Zucman 2016 for the U.S. and Garbinti-Goupille-Piketty 2016, 2017 and Bozio-Garbinti-Goupille-Piketty 2017 for France).

Section 5.1. Imputations for labor income

We usually observe in income tax micro files three different variables for wage income, self-employment income and replacement income (i.e. social insurance income,

including pensions and unemployment benefits). For each of these three categories, we start from the full amount reported on tax returns (before any specific deduction or exemption).

In the pre-tax factor income series, we exclude fiscal replacement income (i.e. we set individual replacement income flows to zero in income tax micro files), and we scale up fiscal wage income and self-employment income flows in order to match the national accounts totals for factor wage income (i.e. employee compensation (D1)) and factor self-employment income (i.e. net mixed income (B3n)) used in the definition of factor income (see table 3). In case no additional information is available, the simplest way to proceed is to apply a simple proportional upgrading rule, i.e. each individual wage income is multiplied by the aggregate ratio between employee compensation and fiscal wage income, and each individual self-employment income is multiplied by the aggregate ratio between net mixed income and fiscal self-employment income. We provide the corresponding computer codes in the country specific studies.

However we usually have more information, i.e. we sometime observe social contributions (in particular pension contributions) on individual tax returns, in which case this information should be used in order to obtain a more accurate imputation. In addition, social security contributions are generally not proportional to wage income or self-employment income: they typically involve lower rates of social contributions for higher labor income brackets (and sometime for lower labor income brackets as well). In the benchmark pre-tax factor income series, we recommend that all available legislative and statistical information on graduated rates of social contributions by wage income and self-employment levels should be used in order to compute factor labor

income and factor self-employment income.[30] Details of imputations procedures are described in specific country study.

In the pre-tax national income series, we scale up fiscal wage income, self-employment income and replacement income in order to match the national accounts totals for pre-tax wage income, self-employment income and replacement income defined on table 4 (broad definition). More precisely, we proceed as follows. We start from factor wage income and factor self-employment income and deduct social contributions using all available information (see discussion above). In case we do not have information on the break-down of social contributions by labor status (i.e. wage earners vs. self-employed workers, a break-down that is not always available in national accounts), then social contributions should be deducted proportionally. We then scale up fiscal replacement income in order to match pretax replacement income (i.e. social insurance income, as defined on table 4).

Section 5.2. Imputations for capital income and wealth

The general idea behind the income capitalization method is to recover the distribution of wealth from the distribution of capital income flows. In its simplest form, the method relies on the assumption of fixed rates of return by asset class (see e.g. Alkinson and Harrison, 1978, and Saez and Zucman, 2016). In more sophisticated versions, one can introduce different rates of return within each asset class, e.g. due to idiosyncratic variations in rates of return, and/or because the rate of return r(k) tends to rise with the

[30] In case social security legislation and social contributions rates vary vastly with employment sector, it might also be necessary to treat differently private sector and public sector employees. This information is typically not available in income tax micro-files and would need to be imputed from other sources. See country-specific studies.

level of asset holding k (see the discussion in section 6 below). More generally, we should make very clear that we certainly do not view the income capitalization method as a magic bullet. Measuring the distribution of wealth is extremely difficult and uncertain, and it is absolutely critical to combine the lessons from the income capitalization method (when available) together with the lessons from other data sources (see section 6 for a detailed discussion). In addition, it should be noted that a number of important asset categories usually do not generate taxable capital income flows (in particular owner-occupied housing), so that it is always necessary to supplement the income capitalization method with household wealth surveys, thereby making it a "mixed method" (using the terminology initially introduced by Atkinson and Harrison, 1978).

In practice, when applying the income capitalization method to income tax micro data, we generally aim to use at least four different categories of assets/liabilities and corresponding capital income flows: housing assets, business assets, financial assets, and financial liabilities (see table 7).

We start with business assets and self-employment income. Note that we do not need to make any assumption about the capital-labor split within self-employment income in order to recover business assets from self-employment income. We simply need to assume that the ratio (self-employment income)/(business assets) is the same, which as a first approximation seems like the most natural assumption. If and when other data sources allow us to do so, we will of course refine this assumption. Note that in some countries available fiscal and national accounts data allow us to split self-employment income and business assets into several subcomponents, so as to refine

the income capitalization method. E.g. in the case of the U.S. we can apply the income capitalization method separately to sole proprietorships, partnerships, and S-corporations (Saez and Zucman, 2016). In the case of France, we observe separately three main types of self-employment income flow,[31] but it is difficult to break down business assets into corresponding categories. Hence, at this stage we apply the income capitalization method with a single category for self-employment income and business assets (Garbinti, Goupille and Piketty, 2016).

We now move to housing assets. We usually observe actual rental income in income tax micro files (i.e. rental income from housing units rented to other households). Using national accounts and estimates of the share of actual rental income in total housing rents (which can usually be estimated using housing surveys), we can scale up actual rental income in a proportional manner. From there we can estimate the value of non-owner-occupied housing by dividing actual rental income by the average of return on housing (as computed on table 7). Imputed rental income (i.e. the rental value of owner-occupied housing) used to be taxable in many countries during the first half of the 20th century (e.g. until 1963 in France). However in most countries it is not taxable anymore, so one cannot observe imputed rental income in income tax declarations (sometime this can be observed indirectly via property tax liability), and we need to use other sources for the imputation of owner-occupied housing (see section 5.2.2 below, where we also address the issue of household debt imputation).

[31] "Bénéfices non commerciaux" for doctors, lawyers, etc.; "bénéfices agricoles" for agricultural income; and "bénéfices industriels et commerciaux" for most other forms of self-employment income.

Finally we come to financial assets. We usually observe a number of different categories of financial asset income in income tax micro-files. There are variations across countries, but generally we observe at least two categories - interest and dividend - which can be scaled up to the corresponding national accounts aggregates in a proportional manner. From there we can estimate interest-bearing assets (currencies, deposits and debt assets) and dividend-bearing assets (equity and investment fund shares), using the categories defined on table 8. The information that is available in income tax micro-files about income attributed to life insurance and pension funds is usually insufficient, so other sources must be used (see section 5.2.2 below). Generally speaking, the information available about financial asset income varies a lot across countries, and we recommend to perform several sensitivity checks regarding the classifications about assets and rates of return that are being used to apply the income capitalization method (see the discussion in section 4.2).

Section 5.2.2. Imputations for owner-occupied housing, pension wealth, debt

The general method used to impute assets that do not generate taxable capital income flows (or assets for which taxable income flows do not provide an adequate indicator to estimate asset holdings) consists of using household wealth surveys. This applies in particular to owner-occupied housing, debt, pension wealth, and other country-specific and legislation-specific financial assets such as life insurance or saving accounts in France (whose return is partly or entirely tax-exempt).

The exact imputation method depends on the characteristics of the available wealth survey. If the survey contains sufficiently many observations, it is better to use a very

flexible imputation method, i.e. one can estimate the percentage of home-owners and average home values for each cell defined by age, gender, percentile of labor income and percentiles of non-housing wealth, and so on for other assets. With smaller surveys, other methods – linear within deciles or quartiles – might be more appropriate. All details and computer codes should in principle be available in country-specific studies (see e.g. Piketty, Saez and Zucman 2016 for the U.S. and Garbinti, Goupille and Piketty 2016, 2017 for France).

Section 5.3. Imputations for taxes and transfers

We now briefly describe the imputation methods that are necessary in order to move from the distribution of pretax income to the distribution of disposable income, following the definitions given on table 5.

For production taxes we recommend simple imputations in proportion to the different pretax national income flows (except for property taxes, which to the extent possible should be imputed in proportion to housing wealth). One may also be tempted to use more sophisticated tax incidence assumptions and associated imputation techniques. E.g. one could assume that production taxes, and in particular value-added taxes, can be imputed partly on factor income, and partly on consumption flows (depending on elasticities of factor supply and demand). However this requires making explicit assumptions about the division of disposable income into consumption and saving at the individual level (see Landais-Piketty-Saez 2011 for an attempt in this direction). One could also take into account different VAT rates and consumption structures by income, etc. Generally speaking we recommend to start with simple imputation

procedures (proportional to pretax national income flows, except for property tax) and to make clear how this could be improved in the future. The benchmark strategy adopted by Piketty, Saez and Zucman (2016) for the US follows an intermediate approach: they assume a simplified profile of saving rates as a function of pre-tax national income, and impute production taxes (except property taxes) in proportion to consumption. In any case, we recommend to compare explicitly the benchmark imputation that is being chosen to an alternative proportional-imputation strategy.

In particular, we should stress that the primary objective behind these imputations is to make income levels (and not only inequality levels) comparable across countries with very different levels of taxation and different tax structures, in particular regarding the relative importance of (indirect) production taxes (which are typically already deducted from fiscal income) and direct income and wealth taxes.

For the corporate tax incidence, the most plausible assumption is that the corporate tax falls not only on corporate equity but also on other forms of financial and business assets as in Harberger (1962)’s seminar analysis as asset owners arbitrage to some extent differences in the net-of-tax returns. We differ from Harberger's analysis only in that we treat residential real estate separately. Because the residential real estate market does not seem perfectly integrated with financial markets, it seems more reasonable to assume that corporate taxes are borne by all capital except residential real estate. We symmetrically assume that residential property taxes only fall on residential real estate. We adopt these tax incidence assumptions for the US and France (see Piketty, Saez and Zucman 2016 and Garbinti, Goupille and Piketty 2016, 2017). In the case of the United States, official distributional estimates of Federal tax

made by the Congressional Budget Office now assume that 25% of corporate taxes fall on labor income (US Congressional Budget Office, 2016). Because US multinational firms can fairly easily avoid US taxes by shifting profits to offshore tax havens without having to change their actual production decisions (e.g., through the manipulation of transfer prices), it does not seem plausible to us that a significant share of the US corporate tax is borne by labor (see Zucman, 2014). By contrast, in small countries---where firms' location decisions may be more elastic---or in countries that tax capital at the source but do not allow firms to easily avoid taxes by artificially shifting profits offshore, it is possible that a sizable fraction of corporate taxes falls on labor. Hence, our assumption on corporate tax incidence should also be seen as provisional, to be potentially modified according to each country's situation, or to be modified as new compelling empirical evidence on corporate tax incidence arises. It is also worth noting that corporate tax incidence assumptions only matter for the distribution of pre-tax income---they do not matter for post-tax series, which by definition subtract all taxes.

Regarding the imputations of personal income and wealth taxes, there are two possibilities that can be used: one can apply legislation over direct taxes and/or use information on tax liability that is often directly observed in income tax micro files. Ideally one may try to combine both approaches. Regarding inheritance taxes (and other personal taxes for which many variables playing a key role in tax computations are typically not available), one can adopt simplified assumptions. E.g. for both France and the U.S. we choose to attribute inheritance tax revenues to top 5% or top 1% wealth holders (in proportion to wealth in excess of the relevant threshold), depending on whether the information available regarding the proportion of decedents and

successors subject to tax and the progressivity of the tax (see Piketty, Saez and Zucman 2016 and Garbinti, Goupille and Piketty 2016, 2017).

Regarding the imputation of cash transfers, one can again combine the use of legislation over cash transfers and the use of information on transfer receipts that is often directly observed in income tax micro files or in the household income surveys (or ideally in the matched income tax/household surveys micro files such as the ERFS surveys in France). In the case of the U.S., we use extensively the information on transfer receipts by income percentile, age and gender extracted from the CPS.

Thanks to these imputations, we can estimate the distribution of pre-tax factor income, pre-tax national income and post-tax disposable income (see tables 3, 4, 5 for the corresponding definitions). In countries with high-quality income tax micro-files and household surveys, these three distributions can be estimated in a relatively precise manner, and in a way that can be compared across countries (a number of imputations assumptions are needed, but one can check by computing variants that the consequences on the series are limited).[32]

The most challenging part is the imputation of in-kind transfers and other public spending (collective expenditures) that is needed to compute post-tax national income (see table 5). Here we should make clear that it is extremely difficult to do this type of

[32] The two items on tables 3 and 4 which require imputation assumptions are government (and non-profit) capital income and surplus of the pension (and social insurance) system. The simplest solution is to attribute government capital income in proportion to pre-tax personal factor income for the computation of pre-tax factor income, and in proportion to pre-tax personal national income for the computation of pre-tax national income. One can also think of more sophisticated rules, such as an imputation 50-50 in proportion to taxes paid and benefits received, and/or different imputation rules for public and private pension surpluses (see Piketty, Saez and Zucman 2016 for a more detailed discussion in the case of the U.S.).

imputation, and it is not even clear whether it really makes any sense to attribute public spending such as roads or police to individuals. The main reason for doing it is to make income levels comparable across countries: otherwise income levels in countries with higher in-kind transfers and collective expenditures would artificially appear to be poorer. This is also the reason for taking them into account in GDP, and this is why we recommend to compute series of post-tax national income. The simplest way (and distribution neutral way) to do this is to attribute all in-kind transfers and collective expenditures in proportion to post-tax disposable income. By doing so, we simply raise all income levels and do not change the distribution.

Another possibility would be to use a lump-sum method : we attribute the same average monetary value of in-kind transfers and collective expenditures to each adult individual. This might be justified for certain in-kind transfers and expenditure, but in some other cases this will vastly overestimate the extent of redistribution. For instance, we observe in most countries highly unequal access to education (children from higher parental income background tend to benefit from higher public education expenditures, particularly because of more extensive access to higher education). Also, unequal life expectancies generate highly unequal access to various public spending. This entails consequences not only for pension receipts (which are not taken into account in our static framework, since pension and unemployment insurance income are already taken into account in pre-tax national income, but which could be included in some future dynamic extension), but also for other public provided services.[33] Also, the value

[33] Note however that as long as we look at cross-sectional inequality, this life expectancy is in effect already taken into account. Consider for example the case of Medicare in the US that provides public health benefits to all individuals aged 65 and above: if the bottom 50% poorest never live beyond 65, then in effect no Medicare health benefits will be attributed to the bottom 50%. However, it is only through a dynamic, generational extension of our inequality series (something that would be desirable, but that is far beyond the scope of these Guidelines at this stage) that such issues could be properly addressed.

of a number of public services - such as police force and protection of property - may rise in proportion to the level of wealth rather than with the level of income.

In the case of the U.S. and France (see Piketty, Saez and Zucman, 2016, and Bozio, Garbinti, Goupille and Piketty, 2017), we provide two sets of post-tax national income series. In our benchmark series, we attribute public health benefits in a lump sum manner (separately for Medicaid and Medicare recipients in the U.S., as identified via CPS, and to all adults in France), and all other in-kind transfers and collective expenditures in a proportional manner. We also provide alternative series with full proportional imputation for all in-kind transfers and collective expenditures.

Section 6. Reconciling wealth inequality sources

In section 5, we described the basic imputation methods used in *WID*.world, which essentially consist of upscaling the labor and capital income flows observed in income tax micro-files in order to match national accounts totals, and of using the income capitalization method in order to recover the distribution of wealth from the distribution of capital income, together with income and wealth household surveys, so as to obtain information on assets that do not generate taxable income flows, as well as on other forms of incomes, taxes and transfers that are not well recorded in income tax micro-files. These basic imputations methods are an attempt to combine in a systematic manner three data sources (national accounts, income tax micro files, and household surveys), using the income and wealth concepts defined in section 3 and the observation units defined in section 4.

We now discuss the limitations of these basic imputation methods, and how they need to refined and reconciled with other data sources that can be used to estimate the distribution of wealth (including fiscal data coming from inheritance taxes and wealth taxes, when they exist), and data on wealth rankings. Generally speaking, we stress that our collective capacity to measure and monitor the distribution of wealth is limited, and that the different data sources at our disposal are not always fully consistent with one another. Our hope is that by combining these data sources in the most explicit manner we can contribute to a better informed public debate on wealth inequality. The perfect data source on wealth does not exist and will never do: one needs to be pragmatic and extract whatever useful information can be extracted from the raw data

sources at hand, as long as this is done very explicitly and by releasing all methodological details and computer codes.

We also stress that the ideal combination of data sources may well vary across countries, partly because different national historical trajectories give rise to different fiscal systems and different data sources. There is nothing new here. In the late 19th and early 20th centuries, British authors were mostly using the income capitalization method to estimate aggregate wealth largely because the schedular income tax system that had been put in place in the mid-19th century in Britain provided regular and reliable estimates on capital income flows. In contrast, French authors favored the estate multiplier method (largely because the availability of extensive inheritance tax data in France, due to the creation of a fairly universal inheritance tax in the late 18th century) (see Piketty 2011 for references). In the U.S., inheritance tax data has always been relatively limited – largely because the federal estate tax created in 1916 provides information solely on the very top of the distribution. Saez and Zucman (2016) have recently shown that the estate multiplier method leads to underestimate the rise of wealth inequality, as compared to the income capitalization method (based upon income tax data, which in the U.S. is relatively high quality) and the SCF wealth survey (also relatively high quality). In contrast, Alvaredo, Atkinson and Morelli (2017) show that British income tax micro-files (and also British national accounts and wealth surveys) make it difficult to apply the income capitalization in a satisfactory manner, and favour the estate multiplier method, largely because inheritance tax data is more comprehensive than in the US. In the case of France, Garbinti, and Goupille and Piketty (2016) show that both the income capitalization and estate multiplier method deliver consistent estimates; they favor the latter for their 1800-1970 wealth distribution

series (as there is no income tax micro file prior to 1970, making it very difficult to apply the income capitalization method, and there is no income tax data at all before 1914), and they favor the income capitalization method for recent decades (post 1970) largely because inheritance tax data has ceased to be annual (and because inheritance tax micro files are not large enough, as opposed to income tax micro-files, which are available since 1970 and offer exhaustive coverage of all income declarations in recent years, like in the US). In sum, there is no perfect data source, and we recommend to use them all and provide a reconciliation between them, to the extent possible in the various countries. We further discuss reconciliation methods below.

Section 6.1. Advantages and limitations of income capitalization method

In theory, the ideal data source to study the distribution of wealth would be high-quality annual administrative data on wealth, based upon automatic transmission of information from financial institutions and real estate transactions to tax authorities. Such data would also be useful for tax authorities in order to properly enforce existing income tax, inheritance tax and property tax legislation (and of course to implement an annual wealth tax). Unfortunately such data usually does not exist for the time being. At this stage, the only source of annual administrative information on wealth generally comes from the income tax (through the annual observation of capital income flows). Inheritance tax data is annual but we observe wealth only at the time of transmission; property tax data usually provides information about real estate only (and it is often based upon economically meaningless cadastral values); wealth tax data generally does not exist, simply because in most countries there is no comprehensive wealth tax

(and when there is one, it often covers a very small fraction of the population and assets).

That being said, there are major limitations with the income tax data and with the income capitalization method. First, there are many countries where a very large fraction of capital income flows is not subject to the progressive income tax any longer and is not reported in income declarations. In particular, interest and dividend income tends to be taxed separately (sometime with a specific tax rate) in a large number of countries, in which case the information on the corresponding income flows often disappears from income tax data. This can severely limit what can be done with the income capitalization method. Next, even in countries where a substantial part of capital income flows are observable in income tax micro-files, we always need to supplement the income tax data with other sources of information (such as wealth surveys) for missing wealth items such as owner-occupied housing or pension funds (see the discussion in section 5 above). Finally, the basic income capitalization method assumes a constant rate of return within each asset class, which may or may not be correct, as we now discuss.

Section 6.2. Reconciling income capitalization and estate multiplier methods

Our objective in WID.world is not to claim that we have discovered perfect data sources and methods to measure income and wealth inequality, but rather to provide plausible and methodical strategies to reconcile the different data sources. One important objective is to reconcile the income capitalization method (which aims to recover the distribution of wealth from the distribution of capital income flows, using income tax

data) and the mortality multiplier method (which aims to recover the distribution of wealth among the living from the distribution of wealth at death, using inheritance tax data). These two methods have long been used by scholars working on inequality, and generally deliver consistent long-run evolutions (see e.g. Atkinson and Harrison (1978), who apply both methods to U.K. income and inheritance tax data ranging from the 1910s-1920s up to the 1970s). However in recent decades the two methods sometime appear to deliver inconsistent results. Using U.S. income and inheritance tax data, Saez and Zucman (2016) found a much bigger rise of top wealth shares with the income capitalization method than with the estate multiplier method (indeed they find very limited or inexistent rise of top wealth shares with the latter method).

There are at least three ways to reconcile the income capitalization and estate multiplier methods, which we note the r(k) bias (differential returns), the m(k) bias (differential mortality), and the e(k) bias (differential tax evasion). First, it could be that the average rate of return to wealth r(k) rises strongly with the level of net wealth k (including within a given asset class), e.g. due to scale economies in portfolio management costs. If this is the case, and if we ignore this, or underestimate the steepness of the r(k) profile, then we will tend to overestimate top wealth shares when we use the income capitalization method (if slightly higher wealth individuals get infinitely higher returns, then one can observe infinite inequality of capital income, even though underlying wealth inequality is relatively small). Next, it could be that the mortality rate m(k) declines strongly with the level of net wealth k. If this is the case, and if we ignore this, or underestimate the steepness of the m(k) profile, then we will tend to underestimate wealth inequality when we use the estate multiplier method (if wealthy individuals never die, then wealth inequality at death will be very small, even

though underlying inequality of wealth among the living is very high). Finally, it could be the relative rate of tax evasion e(k) - i.e. the ability not to report one's wealth to the inheritance tax, relative to the ability not to report one's capital income to the income tax, thanks to legal or illegal reasons - rises with the level of net wealth k.

Assume that the income capitalization and estate multiplier methods deliver different levels of wealth inequality (say, higher top shares with the income capitalization method). It is clear that there are many different combinations of r(k), m(k) and e(k) profiles which can close the gap. To the extent possible, each country-specific study in *WID*.world should attempt to make explicit on what ground one can determine the most plausible combination of r(k), m(k) and e(k) profiles which can reconcile the two methods.

For instance, Saez and Zucman (2016) use external data to estimate the r(k) and m(k) profiles (in particular, data on foundations returns to estimate differential returns, and matched income tax-estate tax data to estimate differential mortality). They find that these two effects are not sufficient to reconcile the two methods. They conclude that the remaining gap is likely to be explained by differential tax evasion, namely a rising fraction of high wealth holdings seems not be reported in the inheritance tax declarations (e.g. because the corresponding assets are located in trust funds that are not subject to estate tax). Again, what is important in these reconciliation attempts is not so much to claim that we are able to measure perfectly well the different effects (which of course we are not), but rather to be as explicit as possible regarding the data sources that we use in order to provide the most plausible reconciliation we can offer, and the potential data sources that could be used in the future to refine the estimates.

Finally, note that an average differential mortality profile m(k) by wealth can arise not only because the wealthy live longer but also because health and longevity can also affect wealth. For instance, it could be that individuals within a given age-wealth cell have private information about their mortality (e.g. there is an onset of a serious sickness). This prior knowledge of death could lead to extra consumption or terminal health spending which could again bias estate multiplier estimates of wealth inequality (particularly if the fraction of population with such prior knowledge has increased over time). As another example, the rate of return r(a) may fall at old age as elderly individuals may loose the ability to manage their finances well, and if this happens sufficiently many years before death, and within asset class, then this can also explain why we tend to underestimate wealth concentration when we use the estate multiplier method.

Section 6.3. Reconciling fiscal sources with wealth surveys and billionaires lists

We also aim to reconcile the fiscal sources (i.e. income tax and inheritance tax data, which can be used to estimate wealth inequality, via the income capitalization and estate multiplier methods, as well as wealth tax data when it exists) with the other available sources on wealth distribution, particularly wealth surveys and billionaires lists. Household wealth surveys are based upon self-reported information, and as such are well known to underestimate top wealth levels. On the other hand they include useful information on certain assets (such as owner occupied housing and pension funds) for which we lack information in existing fiscal sources (see section 5 above). A number of recent studies have tried to supplement household survey data with

billionaire lists - such as those published in Forbes and other magazines - in order to correct upwards the top of the wealth distribution using Pareto interpolation techniques (see e.g. Davies et al (2008, 2011, 2010-2016), Vermeulen (2014), Eckerstorfer et al (2015), Westermeier-Grabka (2015)).

In *WID*.world, we aim to extend and systematize these comparisons and to provide a reconciliation with the wealth inequality estimates based upon fiscal sources (for exploratory comparisons along these lines, see Saez and Zucman 2016, and Piketty 2014, chapter 12). First, we consider that it is important to start from fiscal sources (income and inheritance tax data), which are generally more reliable than surveys and billionaires list (in particular, it is often difficult to know the exact observation unit, wealth concept and methodology used in the billionaire lists). Next, it is not sufficient to estimate a single Pareto coefficient at the top of the wealth distribution: one needs to estimate a generalized Pareto curve, i.e. a curve of Pareto coefficients b(p) varying with the exact percentile (see Blanchet-Fournier-Piketty 2017 and the discussion in section 7 below). E.g. in a country with 100 million adults, billionaires list might be a useful source to estimate the Pareto coefficient at the level of the top 100 or top 1000 wealth holders (i.e. the top 0.0001% or top 0.001%), but will not necessarily be very informative at the level of the top 1 million wealth holders (i.e. the top 1%). It is only by combining the different data sources in a systematic manner that a reconciliation is possible (see the country-specific studies).

We return in section 7 below to the case of countries with limited wealth data and we propose practical and transparent methods in order to exploit whatever data is available (e.g. billionaire data via generalized Pareto interpolation methods). However

we stress that when more systematic data sources exist it is critical to use them in priority and to attempt as much as possible to reconcile the different wealth sources. In particular, in all countries where income tax data, inheritance tax data and wealth surveys are available – at least in part - , it is critical to use them in order to apply reconcile estimates coming from the income capitalization and estate multiplier approaches.

Section 6.4. Reconciling wealth and income inequality series via synthetic savings

Finally, the best way to reconcile the different available wealth sources, and more generally to test the overall consistency of the wealth and income inequality series provided in WID.world, is to analyze explicitly the joint dynamics of the distribution of income and wealth. One critical advantage of the income capitalization approach is that – assuming high quality income tax micro-files are available – one can use the same raw data sources to estimate series on joint distribution of income and wealth. From there one can also estimate series on the wealth-fractile-level and income-fractile-level synthetic saving rates that can account for the transition from wealth and income inequality in time t to wealth inequality in time t+1.

This approach was first developed by Saez and Zucman (2016) in the case of the US and further developed by Garbinti, Goupille-Lebret and Piketty (2016) in the case of France. This approach also forces the authors of the various series to check that the variations in income and wealth inequality trends across countries are consistent with one another.

For instance, Alvaredo, Atkinson and Morelli (2017) find that UK wealth inequality increased somewhat less than in France in recent decades (and much less than in the US), in spite of the fact that top income shares rose more strongly in the UK than in France (and less than in the US). This could be due to smaller inequality of saving rates by income and wealth fractiles in the UK, and/or different fractile-level patterns of portfolio composition, capital gains and rates of return (e.g. higher relative housing price rise in the UK favouring middle class porfolios relative to top wealth shares, even more so than in other countries). As pointed out by Alvaredo, Atkinson and Morelli (2017), housing privatization policies might also have played a role.

We do not have perfect data, so we will not always be able to provide complete answers to these questions. But at least the different available data sources and inequality series should be used in a consistent manner in order to provide as many consistency checks as possible.

Section 7. Countries/years with limited income and wealth data

Until now, we implicitly assumed that we have at our disposal very rich data sources in order to construct “Distributional National Accounts” (DINA) series. Typically, we assumed that we have detailed and reliable national income and wealth accounts

In this section, we discuss the methods that can be used in the case of countries and time periods with more limited data sources, typically with income tax tabulations instead of income tax micro-files, and/or with income tax data covering only a subset of the population rather than the entire population, and/or inadequacy of income tax data (e.g. due to large or complete exemptions for capital incomes). We will take each of this problem in turn, and illustrate the methods that can be used with the case of China (a country with limited access to income tax data; see Piketty, Yang and Zucman, 2017) and France (a country with detailed tax data but with income tax tabulations prior to 1970 rather than micro-files; see Garbinti, Goupille and Piketty, 2017). We will also discuss how initial WTID series using a fiscal income concept can be corrected so as to be more directly comparable to new DINA series. In order to construct WID.world series for countries and time periods with limited data, we strongly recommend to use the “Generalized Pareto interpolation” (gpinter) web interface available on-line (see http://WID.world/gpinter and Blanchet-Fournier-Piketty 2017 for full technical details on Pareto curves and the corresponding interpolation techniques).

Section 7.1. Countries/years with detailed income tax data tabulations

In many countries, particularly in emerging and developing countries, we do not yet have access to income tax micro files, but we do have access to income tabulations. Even in countries where we do have access to micro files (such as the U.S., France or the U.K), such files do not usually exist before 1960 or 1970 (for the most basic micro files), or even before 1990 or 2000 (for the most sophisticated, exhaustive, annual and easily accessible micro files). So for all countries we need to find ways to exploit income tax tabulations, which generally exist since the creation of the progressive income tax (typically between 1880 and 1920 in most countries).

The standard way to exploit income tax tabulations, from Kuznets (1953) to *WTID* series, has been to use Pareto interpolation techniques. There are two limitations with these techniques, which we address in *WID*.world series.

First, observed distributions do not exactly follow Pareto distributions (not even at the top), and this needs to be addressed in a more systematic and rigorous manner than what we did so far. That is, assume that we have at our disposal income tax tabulations indicating the number of taxpayers and total reported income for a number of income tax brackets $[y_1;y_2]$,.., $[y_i;y_{i+1}]$,.., $[y_n;+\infty[$, so that we can compute for each threshold i=1,...,n the inverted Pareto coefficient $b(p_i)$, where p_i is the fraction of the population with income more than y_i, and $b(p_i)$ is the ratio between the average income above y_i and the income threshold y_i. If the distribution of income were truly a Pareto distribution, then $b(p_i)$ should be a constant, at least above a certain threshold. However in practice it is not constant: it is better to think in terms of a "generalized Pareto distributions" characterized by "Pareto curves" b(p) (see figures 3a-3e for France). In previous work, in order to estimate the average income of given decile or percentile, we typically used

the Pareto coefficient that was estimated using the closest income threshold (see e.g. Piketty (2001, appendix B); see also Atkinson (2007) for a discussion of alternative Pareto interpolation techniques).

In *WID.world* series based upon income tax tabulations, the objective is to estimate the entire shape of Pareto curves. Blanchet, Fournier and Piketty (2017) have recently developed the "gpinter" on-line interface (http://WID.world/gptiner) using non-parametric interpolation techniques in order to estimate Pareto curves b(p) (where p = F(y) is the cumulative distribution function) by using a small number of thresholds p_i, and from which one can recover the full distribution F(y). Blanchet, Fournier and Piketty (2017) apply these "generalized Pareto" techniques to large annual micro-files available for the US and France over the 1960-2014 period (a time of rapid changes in the distribution, particularly in the US) in order to test the precision of the method. They show that this leads to more precise estimates for top deciles and percentiles than the standard Pareto-Kuznets-type extrapolations, and most importantly that such techniques can be used to estimate lower deciles, and indeed to generate highly precise synthetic micro-files for the distribution of income for the entire distribution. These techniques can be applied to other countries.

An easy-to-use web interface using these techniques was made available in March 2017 for all users on WID.world/gpinter (users are also offered the possibility to download the R-language computer codes on their own computer). In their simplest version, the programs are designed to transform tabulated data of (y_i,p_i ,b(p_i), i=1,…,n) into g-percentile data, i.e. they compute the quantile function y(p) and Pareto curve b(p) for all 127 g-percentile (see discussion in section 2.2 above and appendix table

A1), together with the corresponding standard errors (which are typically very small if the number of brackets if sufficiently high).

Next, the estimates coming from income tax tabulations need to be corrected in order to homogenize the observation unit and the income concept, which was not done in a systematic manner in existing *WTID* series (see discussion in section 2 above). Regarding the unit of observation, the issue is to deal with countries/years for which the tax unit is the couple rather than the individual. Income tax tabulations usually include information on the fraction of singles and couples within each income bracket. This can be used to estimate separately the Pareto curves b(p) and to generate separate synthetic micro-files for single individuals and for individuals living in couples (under the equal-split assumption), which can then be merged. In order to estimate individual-income inequality series, one would need to make assumptions about the within-couple distribution of income (using estimates available for years with micro-files).

Regarding the income concept, the key issue is to determine how to upgrade fiscal income (as reported in income tax tabulations) into pretax income (as defined using national accounts concepts). Ideally, one would like to apply the same general imputation methods as with micro-files, i.e. one would need to multiply each income category by the corresponding aggregate ratio between pretax income and fiscal income (see section 5 above).[34] By using income tax tabulations and estimating Pareto

[34] There was no systematic attempt to upgrade the different income categories in *WTID* series, except in order to correct for specific fiscal deductions (such as the 10% deduction for "professional expenses" for wages in France, or similar proportional deductions for other income categories, e.g. for dividends in Italy, etc.), as well as a number of general fiscal deductions (e.g. in France the income tax liability of the previous year could be deducted from taxable income). See e.g. Piketty (2001, App. B).

curves, one can estimate the cumulative distribution function for total fiscal income F(y). One would then need to estimate the joint distribution $G(y_l,y_k)$ (where $y=y_l+y_k$=total income, y_l=labor income, y_k=capital income), or more generally the joint distribution with more than two income categories (one may need to distinguish wage income, self-employment income, various forms of rental and financial income flows). In many countries, income tax tabulations provide for each income bracket the breakdown of total income into the various income categories (at least for some years). This could in principle be used to estimate how the capital share $\alpha(y)=y_k/y$ varies with income level y (typically it rises sharply with income level, especially within the top decile and percentile; see e.g. Piketty (2001, Appendix B)). In practice however this strategy turns out to be very difficult to apply, because one needs much more than two income categories in order to properly apply the imputation methods described in section 5.

At this stage we therefore recommend to use simpler strategies. E.g. in the case of France, we have detailed income tax micro files for the 1970-2014 period, so that we can compute the g-percentile-level income ratios between pre-tax national income series and pre-tax fiscal income series. The ratio is larger for higher percentiles (as higher incomes tend to have more tax-exempt capital income flows, typically undistributed profits), and the profile is relatively stable over time (with a steepness rising with the ratio between macroeconomic capital income and fiscal capital income). In our benchmark series we simply apply to pre-1970 g-percentile fiscal income series (coming from generalized Pareto extrapolations) the same average ratios as those observed during the 1970-2014 period. We also show that alternative series (obtained by adjusting the steepness of the ratio profile with the evolution of the macroeconomic-fiscal ratio) are very close to the benchmark series.

In the case of the US, Piketty, Saez, and Zucman (2016) use a simpler methodology to estimate inequality series before 1962 (when no micro-files exist). They rely instead on the top income shares series with income decomposition created by Piketty and Saez (2003). They simply scale up each income component to go from the fiscal income concept to the national income concept. This can create re-ranking issues. Therefore, they do an extra overall adjustment benchmarked on the early 1960s when both the micro-data and the compositional series of Piketty and Saez (2003) are available. Their methodology makes it possible to extend the WTID fiscal income series into WID national income series fairly easily if WTID fiscal income series include income component decompositions and there is at least a few years of micro-data to correct for re-ranking issues.

Section 7.2. Countries/years with limited income tax tabulations

In the previous section we implicitly assumed that income tax tabulations covered the entire distribution, from the bottom to the top of the distribution. This is not always so. In many cases, e.g. in China or India today, we only have income tax data for the top of the distribution, and the question is how to combine it with the household survey data that exists for the lower part of the distribution.

One illustration of how this can be done for the case of China is given by Piketty, Yang and Zucman (2017), who combine household income survey data covering the 1978-2015 period together with recently released fiscal data on high-income taxpayers (those with more than 120 000 yuans in annual taxable income, i.e. roughly the top 1%

or top 0.5% of the distribution) and national accounts data. The most natural assumption is to assume that household surveys provide reliable estimates of the distribution up to a certain percentile (say p=0.9), that income tax data provide reliable estimates – or actually a lower bound, but the point is that it is generally much higher than the survey-based lower bound - above a certain percentile (say p=0.99 or p=0.995, depending on where the tax data starts), and then to make assumptions about the continuous profile of the fiscal/survey upgrade ratio between these two points. In the case of China, we show several variants (linear profile, piecewise linear profile) and find that this makes limited difference. When computing these fiscal/survey upgrade ratios, it is critical to make adequate corrections to the raw fiscal data so as to ensure that it expressed in terms of "fiscal income" (i.e. pre-deductions) rather than "taxable income" (i.e. post-deductions).[35]

Using income tax data to correct survey data usually leads to significant upward corrections of top income shares, as the case of China illustrates (see Piketty, Yang, Zucman 2017). However such a correction is not sufficient: the resulting fiscal income series in order to obtain pre-tax national income series that can be compared to those obtained for countries with income tax micro files. In particular, it is critical to take into account tax-exempt capital income. In practice, important components of capital income are usually missing from fiscal income data, even in the absence of any tax evasion. In particular, the capital income accumulated as undistributed profits of

[35] In practice, raw fiscal data is often expressed in terms of taxable income, so it needs to be adjusted upwards. The most common deductions are deductions for professional expenses (e.g. automatic deduction of 10% for wages in France, up to a ceiling, with no justification), supplementary social contributions (e.g. health insurance premium), and routinely represent 10%-20% of fiscal income (or more), with large variations over time, across countries and across income levels. It should also be noted that deductions can be a powerful way to reduce the tax burden of top income groups (in a less visible manner than reducing rates). They are sometime very creative (e.g. in France during the interwar period it is allowed to deduct the income tax paid on the income of the previous year) and need to be spotted and corrected.

privately owned corporations is usually not included in fiscal income subject to income tax. It is important to correct for this, because the extent to which private shareholders choose to accumulate wealth as undistributed profits (as opposed to distributed dividends and other forms of capital payments such as own-shares buybacks and induced capital gains) may well vary over time and across countries (e.g. due to changing tax incentives), which might introduces various biases in distributional series, particularly at the top of the distribution. As discussed in section 5, the best way to proceed is to use income tax micro files and to upgrade the observed individual-level taxable capital payments (in particular dividends and capital gains) in order to estimate individual-level undistributed profits (using the observed macroeconomic ratio between undistributed profits and dividend payments, and a simple linear upgrading rule, unless other available information suggests otherwise). In order to correct for imputed rent (i.e. owner-occupied housing rent, which is generally not included in taxable income), we recommend to combine income tax micro-files and wealth surveys in order to estimate the distribution of housing wealth and attribute it to the different percentiles of the income distribution. In countries with no access to income tax micro-files, such as China, we need to use other techniques.

One way to proceed (which we apply for instance to the case of China) can be described as follows. First, we estimate from our national accounts series the evolution of total non-fiscal capital income y_{nf}, which we define as the private share of undistributed profits and other tax-exempt capital income flows (including imputed rent) accruing to households. In the case of China, we find that y_{nf} gradually rises from 1% of per adult national income in 1978 to as much as 12% in 2015 (largely due to the rise of private corporate ownership and private housing, and also to the rise of other capital

and business income flows recorded in national accounts and which do not appear in tax data). In contrast, total fiscal income y_f (i.e. total income subject to income tax, before any deduction) represents approximately 70% of national income throughout the 1978-2015 period (had everybody been subject to the income tax). In order to estimate the distribution of total personal income $y_p=y_f+y_{nf}$, we need to make an assumption about the distribution of y_{nf} and the structure of the correlation between y_f and y_{nf}. Regarding the distribution of y_{nf}, we assume it follows the same distribution as the distribution of wealth, which we estimate by applying generalized Pareto interpolation techniques to household wealth surveys and wealth rankings. Regarding the correlation structure between y_f and y_{nf}, on the basis of estimates done for countries with adequate micro-files (in particular the U.S. and France), one can use the family of Gumbel copulas, with a benchmark Gumbel parameter around $\theta=3$ (see Blanchet, Fournier and Piketty, 2017).[36] In practice, one can show that assuming Gumbel parameters in the 2.5-5 range instead of 3 has a relatively small impact on the final series (see Piketty, Yang and Zucman 2017).

Section 7.3. Countries with no income tax data

There are also countries for which we have no access at all to income tax data (not even the kind of tabulated data for top incomes that we have for China). In case we only have access to household survey data (micro-files or tabulated data), there are two ways to proceed. One way is to try to correct the top of the survey-based distribution (which typically involves very low inverted Pareto coefficient b(p) above

[36] Gumbel parameter $\theta=1$ corresponds to perfect independence, and $\theta=+\infty$ to perfect correlation. Observed distributions for the U.S. and France over the 1960-2014 period are well approximated by Gumbel parameters around 2.5-3.5.

p=0.9, sometime very close to one in case of a near complete absence of top incomes in survey respondents and/or the use survey top coding) by applying more plausible Pareto coefficients, on the basis of what we observe in similar countries where income tax data is available. In their study of the world distribution of income, Lakner and Milanovic (2015) and Anand and Segal (2015) use a somewhat similar method (in spirit): they run a regression between WTID top income shares and survey-based inequality measures (decile shares) in countries where both are available, and use the regression coefficients to correct top income shares in non-WTID countries (i.e. countries with no income tax data) (see also Chancel and Piketty, 2015 for a discussion of these methods). One could also make this method more systematic by using the notion of Pareto curves and applying the generalized extrapolation techniques developed by Blanchet, Fournier and Piketty (2017).

The other (and potentially complementary) way to proceed is to use a method that is similar in spirit to the copula method described above. That is, assume that available survey data gives us reliable information on the distribution of labor income y_l (possibly with a need for a correction at the top) and no reliable information at all for the distribution of capital income y_k. Then one can estimate independently the distribution of capital income y_k from information on the distribution of the wealth (coming from wealth surveys and wealth rankings), and then estimate the distribution of $y=y_l+y_k$ by using the family of Gumbel copulas, typically with a benchmark Gumbel parameter around $\theta=3$ (see Blanchet, Fournier and Piketty, 2017). Additional work needs to be done in order to test the robustness of these methods.

Section 7.4. Transforming WTID series into DINA series

At this stage, most top income shares series in WID.world are still expressed in terms of fiscal income. A key issue for the future is how to transform them into series that can be compared to the income shares series expressed in terms of pre-tax national income. There are two issues here. One is to correct for the observation unit, e.g. to transform tax-unit-based top fiscal income series into equal-split top fiscal income series. In case we have information on the fraction of singles (vs married) in income brackets (which is usually available in raw income tax tabulations), then one can apply the generalized Pareto extrapolation routines. The other possibility is to use the g-percentile-level equal-split/tax-unit income ratios observed for countries and years for which both are available. The other issue is to correct for the income concept. Again there are two possibilities. The best one is to estimate the amount of missing capital income (in particular undistributed profits) and to apply the Gumbel copula techniques described above. One can apply g-percentile-level pre-tax-national-income/fiscal-income ratios observed for countries and years for which both are available. Again, more work is needed in order to test the robustness of these methods, and future versions of these Guidelines will reflect this.

Also equal-split adults vs tax units vs individualistic adults

Section 7.5. Working with limited tabulated data and the WID.world/gpinter interface

In order to construct WID.world series for countries and time periods with limited data, we strongly recommend to use the “Generalized Pareto interpolation” (gpinter) web interface available on-line (see http://WID.world/gpinter and Blanchet-Fournier-Piketty

2017 for full technical details on Pareto curves and the corresponding interpolation techniques). We should stress however the quality of the g-percentile output series (or synthetic output micro-files) delivered by the web interface ultimately depend upon the quality of the tabulated input data. As shown by Blanchet-Fournier-Piketty 2017, two conditions are critical to guarantee the quality of the output.

First, the tabulated input data does not necessarily need to include a large number of brackets, but there must be at least some brackets covering the entire distribution, e.g. with thresholds around p=0.1, p=0.5 and p=0.9. In case the tabulated data does not includes at least one threshold around p=0.1 or p=0.3, then it is going to be very difficult for gpinter to properly interpolate the shape of the bottom half of the distribution. Similarly, in case the tabulated does not include at least one threshold around p=0.9, then it is going to be very difficult to properly interpolate the share of the top decile of the distribution. Blanchet-Piketty-Fournier provide bounds of estimation errors depending on the number and location of the thresholds of the input tabulated data.

Next, the precision of the method depends critically on the fact that the tabulated data includes information on the number of individuals between any two thresholds as well as on the average income or wealth (and/or total income or wealth) between any two thresholds. This is the essence of the Pareto curve interpolation method: one needs both pieces of information in order to compute the b(p) coefficients.

In case these two conditions are met, say if one uses tabulated data with at least one threshold around p=0.1, one threshold around p=0.5 and one threshold around p=0.9, including reliable information of the numbers of individuals and their average or total

income or wealth in all brackets, then the gpinter interface will generate very reliable results (with errors typically less than 0.1%; see Blanchet-Fournier-Piketty 2017 for more details).

In case part of this information is missing, then the WID.world/gpinter interface will still deliver some results (under certain conditions; see below), but they will be less precise. For instance, in case the user solely provides bracket-level information on averages for income or wealth (i.e. information on fractiles *p* and on shares or averages using "*bracketsh*", "*topsh*", "*bracketavg*" or "*topavg*", using gpinter notations), and no bracket-level information on thresholds ("*thr*"), the interface will still provide output tables by interpolating the Lorenz curve. However we stress that shares-based interpolation is bound to be less precise than thresholds-and-shares-based generalized Pareto interpolation, and we urge users to be cautious. In order to improve accuracy, we offer the possibility to specify the inverted Pareto coefficient b for the top bracket (e.g. on the basis of observed b for countries and years with similar Lorenz curves). In case this information is provided by the user it will be exploited by the interpolation procedure (otherwise it will be estimated from Lorenz curve interpolation).

Similarly, in case the user solely provides bracket-level information on thresholds (ie. information on fractiles *p* and matching quantiles "*thr*"), and no bracket-level information on averages for income or wealth (i.e. no information on shares or averages using "*bracketsh*", "*topsh*", "*bracketavg*", "*topavg*" or "*b*"), the interface will provide output tables by interpolating the quantile function. In case the user also provides information on the overall average income or wealth for the entire population this information will be exploited by the interpolation procedure (otherwise it will be estimated from quantile function interpolation).

We stress that thresholds-based interpolation is particularly fragile. It is much less precise than thresholds-and-shares-based generalized Pareto interpolation, and we urge users to be cautious. In order to raise accuracy, we offer the possibility to specify the inverted Pareto coefficient b for the top bracket (e.g. on the basis of observed b for countries and years with similar quantile functions below the top). In case this information is provided by the user it will be exploited by the interpolation procedure (otherwise it will be estimated from quantile function interpolation). We strongly recommend to use this option. Thresholds-based interpolation is particularly fragile at the very top: small variations in the exact location of the top two thresholds and on small errors in the raw data can generate potentially large variations in interpolated Pareto coefficient for the top of the distribution, and hence large variations in top shares and standard inequality indicators). It is also very fragile at the very bottom: small variations in the input data can generate large variations in the interpolated mass of the population of zero or near-zero income or wealth, again with potentially large consequences on shares and inequality indicators. When making comparisons over time and countries, one needs to ensure that differences in output series are not driven by differences in the form of the input data.

If used with caution, however, threshold-based interpolation (and also sometime share-based interpolation) can be very useful in order to exploit threshold-based historical data (see e.g. Novokmet, Piketty and Zucman (2017) for the case of threshold-based historical data for the Soviet Union).

Section 7.6. Countries/years with limited aggregate wealth data

Worst case scenario: no balance sheet to estimate aggregate wealth, no household wealth survey. Quite common. Nevertheless, one can use billionaire data. See e.g. Novokmet-Piketty-Zucman 2017 on Russia or Alvaredo-Assouad-Piketty 2017 on Middle East.

First step: aggregate wealth estimate. When balance sheets not available they can be constructed using census-type methods. See e.g. China, Russia, Greece… Otherwise, one can also use WB (2006, 2011) wealth estimates. Finally, if no reliable information at all is available, one can attribute beta 400% (300-500% variants) or other average values. The relation between beta and development is complex, so better to take simple benchmark values rather than imputation methods of the kind used by Davies et al (2008, 2011, 2010-2016).

Notes on WB wealth estimates. World Bank 2006, p.144-145 on pure rate of time preference of 1.5% over 25 years (g effect neutralized by assuming gama=1) (same in WB 2011 p.142-143 ; total wealth estimated from total consumption, intangible wealth (including human) as residual from produced wealth (PIM) and natural wealth (again NPV); no use at all of balance sheets or market values for housing or equity; see also WP 2011 p.94-95 on theta=1.5% and gamma=1; this all comes from Hamilton-Hartwick 2005) (interesting, but quite arbitrary flow multiplier and totals) (same pb with direct human K attempt in WB 2011 p.108: btw 7 and 11 GDP)
(they note WB 2006 p.125 that SNA/SEEA guidelines mention both the NPV method and the net stock price method for evaluating natural resources, and then to favor NPV, but its unclear why) (if we were to the same for housing rent it would deliver strange value)

See WP 2011 p.129-138 for survey of current country practice (p.134: Indonesia and Mexico use net stock price method for natural ressources)

Pb WB estimates….

Section 7.6. Countries/years with limited wealth distribution data

Second step: using billionaire data to interpolate wealth inequality. Simple method using billionaire share in aggregate national income and private wealth and generalized Pareto interpolation methods; see NPZ 2017 and AAP 2017 for explanations. Similar in spirit to Davies et al (2008, 2011, 2010-2016), but more transparent and explicit. More sophisticated method using individual-level data on billionaires: see Blanchet (2016); but requires individual-level survey data; and uncertainty so large that reduced-form method preferable (see Blanchet 2016 revised version).

Section 8. Concluding comments

In these Guidelines, we have described the basic concepts, sources and methods that we in the World Inequality Database (WID.world). We should stress again that these methods are fragile, exploratory and subject to revision. As more countries join the database, new lessons will be learned, and the methods will be refined and updated. Accordingly, new updated versions of these Guidelines will be regularly released on WID.world.

References

R. Aaberge, T. Atkinson, S. Königs, C. Lakner, "From Classes to Copulas: Wages, Capital and Top Incomes", 2013

F. Alvaredo, T. Atkinson, S. Morelli, "Measuring Wealth Inequality in the U.K.", PSE and Oxford, 2017

A. Atkinson, "Top Incomes in the U.K. over the 20*th* Century", *Journal of the Royal Statistical Society*, 2005

A. Atkinson, "Measuring Top Incomes: Methodological Issues", in *Top Incomes over the Twentieth Century....*, chap.2, p.18-43, Oxford: Oxford University Press, 2007

A. Atkinson, A. Harrison, *Distribution of Personal Wealth in Britain*, Cambridge University Press, 1978

A. Atkinson, T. Piketty (eds.), *Top Incomes over the Twentieth Century: A Contrast between Continental European and English-Speaking Countries* (Volume 1). Oxford: Oxford University Press, 2007

A. Atkinson, T. Piketty (eds.), *Top Incomes : A Global Perspective* (Volume 2). Oxford: Oxford University Press, 2010

A. Atkinson, T. Piketty, E. Saez, "Top Incomes in the Long Run of History", *Journal of Economic Literature*, 2011, 49(1), pp. 3-71.

T. Blanchet, L. Chancel, “National Accounts Series Methodology (WID.world)”, WID Methodological Notes, 2016

T. Blanchet, J. Fournier, T. Piketty, "Generalized Pareto Curves: Theory and Applications", WID.world Working Paper 2017/03

A. Bozio, B. Garbinti, J. Goupille, T. Piketty, "Inequality and Redistribution in France, 1900-2014: Evidence from Post-Tax Distributional National Accounts (DINA)", PSE, 2017, in progress

L. Chancel, A. Gethin (2017) "Building a global income distribution brick by brick" WID.world Technical Note 2017/5

J. Davies (ed.), *Personal Wealth from a Global Perspective*, UNU-Wider Studies in Development Economics, Oxford University Press 2008

J. Davies, S. Sandstrom, A. Shorrocks, E. Wolff, “The Level and Distribution of Global Household Wealth”, *The Economic Journal*, 2011

J. Davies, R. Lluberas, A. Shorrocks, *Global Wealth Report and Databook*, Credit Suisse Research Institute, 2010-2016, annual publication

P. Eckerstorfer, J. Halak, J. Kapeller, B. Schutz, F. Springholz, R. Wildauer, "Correcting for the Missing Rich: An Application to Wealth Survey Data", *Review of Income and Wealth*, 2015

B. Garbinti, J. Goupille, T. Piketty, "Income Inequality in France, 1900-2014: Evidence from Distributional National Accounts (DINA)", WID.world Working Paper 2017/04

B. Garbinti, J. Goupille, T. Piketty, "Accounting for Wealth Inequality Dynamics: Methods, Estimates and Simulations for France (1800-2014)", WID.world Working Paper 2016/05

Hamilton, K., J. Hartwick, “Investing Exhaustible Ressource Rents and the Path of Consumption”, *Canadian Journal of Economics*, 2005

Harberger, Arnold C. “The Incidence of the Corporation Income Tax,” Journal of Political Economy 70(3), 215--240.

S. Kuznets, *Shares of Upper Income Groups in Income and Savings*, New York: National Bureau of Economic Research, 1953

C. Lakner, B. Milanovic, "Global Income Distribution: from the Fall of Berlin Wall to the Great Recession", World Bank, 2013

C. Landais, T. Piketty, E. Saez, *Pour une révolution fiscale - Un impôt sur le revenu pour le 21e siècle*, Seuil, 2011 (www.revolution-fiscale.fr)

T. Piketty, *Les hauts revenus en France au 20e siècle - Inégalités et redistributions, 1901-1998*, Grasset, 2001

T. Piketty, "Income Inequality in France, 1901-1998", *Journal of Political Economy*, 2003

T. Piketty, “On the Long-run Evolution of Inheritance: France 1820-2050”, Quarterly Journal of Economics, 2011

T. Piketty, E. Saez, "Income Inequality in the U.S., 1913-1998", *Quarterly Journal of Economics*, 2003

T. Piketty, E. Saez, G. Zucman, "Distributional National Accounts: Methods and Estimates for the U.S.", NBER Working Paper No. 22945, December 2016.

T. Piketty, L. Yang, G. Zucman, “Capital Accumulation, Private Property and Rising Inequality in China”, WID.world Working Paper 2007/06

T. Piketty, G. Zucman, "Capital is Back: Wealth-Income Ratios in Rich Countries 1700-2010", *Quarterly Journal of Economics*, 2014

F. Novokmet, T. Piketty, G. Zucman, “From Soviets to Oligarchs: Inequality and Property in Russia 1905-2016”, WID.world Working Paper 2007

E. Saez, G. Zucman, "Wealth Inequality in the U.S. since 1913: Evidence from Capitalized Income Tax Data", *Quarterly Journal of Economics*, 2016

US Congressional Budget Office. “The Distribution of Household Income and Federal Taxes, 2013”, US Congressional Budget Office report, Washington DC, 2016.

Vermeulen, P., "How Fat is the Top Tail of the Wealth Distribution ?", ECB, 2014

Westermeier, C., M. Grabka, "Significant Statistical Uncertainty over Share of High Net Worth Households", DIW Economic Bulletin, 2015

World Bank, *Where is the Wealth of Nations? Measuring Capital for the 21st Century*, 2006

World Bank, *The Changing Wealth of Nations. Measuring Sustainable Development in the New Millennium*, 2011

Zucman, G., “The Missing Wealth of Nations”, Quarterly Journal of Economics, 2013

Zucman G., “Taxing Across Borders”, Journal of Economic Perspectives, 2014

Appendix. List of supplementary documents and material

DINAGuideslines_Income_Wealth_Concepts_WID.xlsx

DINAGuideslines_Income_Wealth_Concepts_WID.pdf