DISTRIBUTIONAL NATIONAL ACCOUNTS GUIDELINES

Methods and Concepts used in the World Inequality Database

VERSION: OCTOBER 2025

WORLD INEQUALITY LAB

Distributional National Accounts Guidelines

Methods and Concepts Used in the World Inequality Database

This version: October 22, 2025

Coordinators

Lucas Chancel
Ignacio Flores
Rowaida Moshrif
Gastón Nievas
Thomas Piketty

Contributors

Facundo Alvaredo
Anthony B. Atkinson †
Luis Bauluz
Thomas Blanchet
Pierre Brassac
Jonas Dietrich
Matthew Fisher-Post
Valentina Gabrielli
Bertrand Garbinti
Amory Gethin
Jonathan Goupille-Lebret
Clara Martínez-Toledano
Marc Morgan
Theresa Neef
Anne-Sophie Robilliard
Emmanuel Saez
Li Yang
Gabriel Zucman

† *Honorary Member, In Memoriam*

This is an updated and extended version of the DINA Guidelines first published in 2016, then revised in 2020[1].

[1] The document dated 2024 corresponds to the 2020 Guidelines with only minor editorial updates, such as the addition of references to statistical packages.

Available Tools and Supporting Resources

We encourage readers to explore the following resources, available online:

- Codes Dictionary Page: A key reference for understanding the structure of the WID database and the construction of its variables.
- Gpinter Tool: Designed for users interested in estimating income and wealth distributions from raw data, the `gpinter` tool is accessible both via a web interface and as an R package.
- Publication Library: Browse research papers, working papers, and reports based on the `WID.world` data.
- Technical Notes: Detailed documentation describing concepts, methods, and data processing choices used to construct the database.
- Replication Packages: Direct links to datasets and code used to replicate key results published by the World Inequality Lab.

Data Download Options: Depending on your needs, there are two main ways to access the data:

- Direct Download: Our user-friendly data page allows you to browse and download datasets directly in the format that best suits your work. Use the left-hand menu to select the data you need.
- Statistical software: For more technical users, data can also be accessed through STATA and R. Comprehensive documentation is provided to guide you through these methods.

To ensure full transparency, the World Inequality Lab makes available the full set of `WID.world` computer codes used to construct the database from both researchers' data and official sources. These codes are regularly updated and accessible via GitHub.

All these resources are designed to make the `WID.world` database as accessible and useful as possible for users of all backgrounds. For further general questions, feel free to contact us at `info@wid.world`, for specific questions about our statistical packages, please use `stats@wid.world`.

Contents

Introduction

These DINA Guidelines describe the concepts, data sources, and methods used in the WID.[2] After their initial publication in 2016 and major revision in 2020, the database has expanded quickly, with new variables, improved estimates, and wider coverage. This third edition is an effort to consolidate that progress, introducing revised macroeconomic aggregates, updated income and wealth definitions, enhanced data-quality flags that capture changes over time and across series, and new sections on wealth distribution and gender inequality. Much of this work had previously been reported in separate papers and technical notes; the present edition centralizes and explains these developments. Before reviewing the content, we begin with a short presentation of the WID and its history.

The World Inequality Database and Its History

The WID is an extensive database on the distribution of income and wealth, both between and within countries. It is primarily maintained by the WIL, located at the PSE. But it is fundamentally the result of a coordinated, collaborative effort involving hundreds of researchers throughout the world over the past twenty years.

In recent years, the WIL has been leading efforts in creating "distributional national accounts." The goal is to provide estimates of the distribution of income and wealth that are harmonized over time and across countries, that are consistent with the macroeconomic aggregates produced by national statistical institutes, and that can therefore be viewed as a distributional extension of the existing international SNA.

The motivation behind this project runs deep. The starting point of the WID was the work of Piketty (2001, 2003) and Piketty and Saez (2003) on the long-run evolution of top incomes in France and in the United States. They used historical tax records to construct estimates of

[2] See https://wid.world.

income inequality since the early 20th century. Unlike surveys — which were and in many cases still are the primary source used to study inequality — tax data can cover very long time frames, and can accurately capture the very top of the distribution.

The use of tax data to study inequality is not new. In fact, it is part of a long tradition. Kuznets (1953) pioneered the use of tax data to measure top income shares, and Atkinson and Harrison (1978) used them to study wealth inequality. The work of Piketty (2001, 2003) and Piketty and Saez (2003) was largely inspired by these contributions. But in the past twenty years, a flourishing literature on the topic has emerged. In particular, by combining historical tax data and national accounts in a systematic manner, a succession of studies constructed top income share series for a large number of countries. These projects generated a large volume of data, intended as a research resource for further analysis, as well as a source to inform public debate on income inequality.[3]

The WTID was created in January 2011 to provide easy access to all these series. It eventually included relatively homogeneous series on income inequality for more than 30 countries, spanning most of the 20th and early 21st centuries. More than one hundred researchers from all parts of the world have contributed to the WTID. These series established important new facts about the long-run evolution of inequality, and had a large impact on the global inequality debate.

Nevertheless, the WTID had many limitations. The sole reliance on tax data meant that we were dependent on the local legislation of countries in any given year when it came to the income concept or the statistical unit under consideration. Because of exemption thresholds, we generally could not provide data beyond the top 10% of the population — or in some cases even narrower groups. The database was limited to pretax income, and had no information on redistribution, or on wealth. And many methodological aspects had not been harmonized. To overcome these limitations, the WTID needed to undergo major changes.

The first of these changes occurred in December 2015, when the WTID was subsumed into the WID. In addition to the WTID top income shares series, this first version of the WID included an extended version of the historical database on the long-run evolution of national wealth and national income that was first developed by Piketty and Zucman (2014).[4] We changed the name of the database from WTID to WID in order to express the extended scope and ambition of the

[3]See the two multi-country volumes on top incomes edited by Atkinson and Piketty (2007, 2010). See Alvaredo, Atkinson, Piketty, et al. (2013) and Atkinson, Piketty, and Saez (2011) for surveys of this literature.

[4]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.

database and the new emphasis on both wealth and income.

Meanwhile, we started work on the harmonization and the extension of the top income share series available in the WTID. For that, we needed a clear framework for how to conceptualize both income and wealth. Rather than creating such a framework from scratch, we opted for one that was already in place, widely used and accepted by most countries: the SNA. Two pilot projects, one in the United States (Piketty, Saez, and Zucman, 2018; Saez and Zucman, 2016) and one in France (Bozio et al., 2020; Garbinti, Goupille-Lebret, and Piketty, 2018, 2020) created the first distributional national accounts and included consistent estimates for pretax income, post-tax income and wealth. Other countries started to follow, including some for which the available data was much more limited (e.g. Piketty, Yang, and Zucman, 2019). We synthesized our experience from this new research in the first version of the DINA guidelines. We also created a new website to match the extended scale and ambition of the WID. It was made public at `https://wid.world` in January 2017, and included new data visualization tools.

Towards a Comprehensive Database on Income and Wealth

The evolution from the former WTID to the current version of the WID has led to an extension of the database in several directions. First, we have started to cover more and more countries, including from the developing world. The former WTID was limited to a very specific type of data, and as a result it remained largely centered around Western countries. The new WID seeks to achieve universal coverage, by developing methodological standards for dealing with countries with limited data.[5] The former WTID focused on tax data and dismissed surveys almost entirely. But in fact surveys provide information that is essential and complementary to tax data. Thus, in order to get the best estimates the WID recognizes the need to combine the various sources. The WTID only covered the top 10% of the population, while the WID covers the full distribution from top to bottom. The WTID focused on pretax income inequality, while the WID provides extensive data on the distribution of pretax income, post-tax income and wealth, on top of a detailed decomposition of macroeconomic aggregates of income and wealth.

The dual focus on income and wealth aggregates *and* on their distribution is a key feature of the WID. We hope this will contribute to reconcile inequality measurement and national accounting. In some cases this may require revising some national accounts concepts and estimates, and we

[5] We are currently on track to provide recent income inequality estimates for nearly every country in the world, and plan to ultimately do the same for wealth.

believe that both inequality measurement and national accounting have something to gain from this endeavor. By combining the macro and micro dimensions of economic measurement, we are, of course, following a very long tradition. The first national accounts in history — King's famous social tables, produced in the late 17th century — were in fact distributional national accounts, showing the distribution of England's income, consumption, and saving across 26 social classes — from temporal lords and baronets down to vagrants — in 1688 (see Barnett, 1936). Simon Kuznets was one of the founders of national accounting, and also the first scholar to combine national income series and income tax data in order to estimate the evolution of top income shares 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 not only income but also wealth and its distribution — 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 the wealth distribution in Britain over 1922–1972.

Needless to say, such an ambitious long-term objective — annual distributional national accounts for both income and wealth for all countries in the world — will require a very broad international and institutional partnership. We certainly do not claim that the WID project in its current form has the capability to achieve this objective alone. It started as an informal academic network. Over the past fifteen years, it has been financed by a number of research grants from public research agencies — in particular the ERC — and non-profit institutions.[6] It will keep evolving in the future, and in order to achieve its long-run objective, new partnerships will undoubtedly need to be developed, and are currently being 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.

Indeed, the work of other organizations in promoting methodologies to compile distributional results in line with macroeconomic aggregates has been flourishing in recent years. Notable initiatives include the OECD-Eurostat expert group on disparities in a national accounts framework (EG DNA) and the ECB expert group on distributional financial accounts (EG DFA). These groups have focused their attention on income, consumption and savings (EG DNA), as well as on wealth (EG DFA), with member countries already engaged in providing experimental results according generic guidelines (see Zwijnenburg, Bournot, and Giovannelli (2017), OECD (2019), Törmälehto (2019), OECD (2020), and ECB (2019)). While similar in spirit to the methodology described in these guidelines, our approach diverges in some important ways. Principally among them is the fact that the sole focus of these organizations is to distribute the disposable income

[6]See https://wid.world/funding-partners/.

of the household sector in the SNA, whereas the DINA we envisage covers the entire economy. We seek to distribute the entirety of national income among resident households (including all income flowing to corporations, the government, and to and from the foreign sector). In this way we account for 100% of macroeconomic growth coming from GDP statistics. We also present results for numerous concepts (e.g. pre-tax national income, post-tax disposable income and post-tax national income) across granular percentile groups reaching small fractiles at the very top of the distribution, with greater precision than the aforementioned studies. Our motivation for doing so are described in the chapters that follow.

Goal and Content of the Guidelines

Given the scale and ambition of the WID project, we feel that it is important to set standards, clarify the concepts and the methods, and provide guidance for future applications to new countries. Hence the present guidelines. Like the SNA itself, these guidelines are provisional and subject to revisions — and always will be. We should stress at the onset that our methods and series are and will always be imperfect. Nevertheless, we attempt to combine the different types of data that are available 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.

Additional details are provided in the research papers developing DINA estimates for specific countries (see in particular Piketty, Saez, and Zucman (2018) for the United States; Bozio et al. (2020) and Garbinti, Goupille-Lebret, and Piketty (2018, 2020) for France; Alvaredo, Atkinson, and Morelli (2018) for the United Kingdom; Piketty, Yang, and Zucman (2019) for China). More information is also available in the regional DINA papers: see Blanchet, Chancel, and Gethin (2022) for Europe; Chancel, Cogneau, et al. (2019) for Africa; Jenmana, Yang, and Khaled (2020) for Asia, and De Rosa, Flores, and Morgan (2024) for Latin America.

We also refer all interested readers to a number of research papers and technical notes documenting the extension of the World Inequality Database to more extended historical series and a larger set of economic variables. See in particular Bajard, Bauluz, et al. (2025), Bauluz, Brassac, et al. (2024), Chancel and Piketty (2021), Gomez-Carrera et al. (2025), Nievas and Piketty (2025b), and Nievas and Piketty (2024). The main methodological takeaways are summarized in the present Guidelines, but readers who are interested in additional details should consult the full collection of our working papers and technical notes

The content of these Guidelines is structured around a clear separation between the concepts (in part one) and the methods (in part two). The first deals with the theoretical conceptualization of income and wealth, the definition of the different subcomponents of income and wealth, and how they should be distributed to individuals. The second part focuses on practical methods, on how to deal with the different types of data.

Chapter 1 covers general issues, such as the definition of the statistical unit, price indexes, and currency conversion factors. Chapters 2 and 3 set out the conceptual definitions of income and wealth. In this edition, Chapter 2 has been expanded with new material on the international consistency of income flows, revisions to post-tax income definitions, and a new conceptual section on gender inequality. Chapter 3 now documents methodological adjustments to the international consistency of foreign asset positions. Chapter 4 explains how aggregate income and wealth series are constructed and has been revised to update its description of wealth aggregates and of the balance of payments. Chapters 5 and 6 describe the estimation of DINA series for income and wealth in countries with relatively good data coverage; in this edition, Chapter 5 has been updated with a revised treatment of government spending, taxes, and benefits, and now documents the estimation of the female labor income share. Chapter 6 now adds new documentation on global wealth distribution imputations. Chapter 7 addresses estimation strategies for countries with much more limited data, and Chapter 8 discusses how global inequality estimates are produced from the database. Finally, Chapter 9 presents indicators of data quality and transparency, which have been refined in this edition to incorporate time variation and provide series-specific assessments beyond distributional income.

Part I

Concepts

Chapter 1

Measurement Concepts

This chapter focuses on the conventions used by the WID with respect to the unit of observation, price indexes, currency conversion factors, and indicators. The conventions used with respect to the definitions of income and wealth are explained in chapters 2 and 3.

1.1 Unit of Observation

One of the major limitations of inequality estimates based on raw tax data, such as those published in the former WTID, is the lack of homogeneity in the unit of observation. Most WTID series were constructed using the "tax unit" as defined by the tax law of the country at any given point in time. In joint taxation countries like France or the United States, the tax unit has been defined as the married couple (for married individuals) or the single adult (for unmarried individuals), and the top income shares series that used to be produced for these two countries (Piketty, 2001, 2003; Piketty and Saez, 2003) did 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).[1]

To ignore tax unit composition is problematic in measuring inequality, as variations in the share of single individuals (or in the extent of assortative mating in couples) along the income distribution of tax units can affect the evolution of income inequality in confounding and contradictory ways. Furthermore, in some countries, the tax system switched from joint to

[1] That is, 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.

individual taxation (e.g., 1990 in the United Kingdom), which creates additional comparability problems in the WTID series (see Atkinson, 2005, 2007). The WTID series also used to distribute income among different population age groups depending on the country (16 and older, 18 and older, or 20 and older), which creates similar problems.

To address these issues, the DINA series strive to use homogeneous concepts for the unit of observation and the overall population. The two key questions concern the population to which we distribute income (adults only or full population), and how to split income within a couple or a household. Our benchmark series consist of "equal-split adults" (income distributed to adults and distributed equally within couples or households). To the extent possible, we also aim to produce "individualistic adult" series that distribute income specifically to each individual, and "full-population" series that distribute income to the whole population (including children).

We use "equal-split adults" as our benchmark both for conceptual and data availability reasons. These series are meaningful when it comes to studying the income distribution, and also relatively easy to compute in a comparable way across countries. Yet we stress that all the different conventions are valuable and provide complementary information. Because income is mostly earned by adults, it makes sense to distribute it only to adults when it comes to studying how much different people earn, which is our primary goal. But it also makes sense to distribute it across the whole population (including children) to study the distribution of how much people can consume, which can be a better proxy for standards of living. How to distribute incomes within a couple or a household also offers interesting and complementary perspectives on different dimensions of inequality. The equal-split perspective takes the view that couples redistribute income and wealth equally between its members. This is arguably a very optimistic — even naive — perspective on what couples actually do: bargaining power is typically very unequal within couples, partly because the two members come with unequal income or wealth. But the opposite perspective (zero sharing of resources) is not realistic either, and tends to underestimate the resources available to nonworking spouses (and therefore to overestimate inequality in societies with low female participation in the labor market).

1.1.1 Total and Adult Resident Population

The DINA series should account for the distribution of the income and wealth of a country among its entire resident population. The definition of the resident population follows the definition of the SNA, which is itself based on the definition given by the IMF in its BPM6 (IMF, 2009). Resident households of a given country include any household for which the country constitutes its “predominant center of economic interest.” This very extensive definition ensures that any person in the world belongs to one — and only one — economic territory. Therefore, this definition is wider than the population under consideration in most surveys. Surveys typically exclude from their sampling frame people living in collective facilities such as hospitals, retirement homes, student accommodations, convents, or prisons. While these only represent only a minority of people, they can be overrepresented at the bottom of the distribution. To the extent possible, DINA series should make the effort to include them.

Within the resident population, our benchmark DINA series consider the subpopulation of adults, by which we mean people of age 20 and older. The main motivation behind this restriction is that adults are the primary earners of income, so that it makes more sense to restrict ourselves to them.[2] We refer to this convention as “per adult” wealth or income. It could also be justified to attribute the income of a household to all its members, including children, if we specifically wish to study the contribution of income to the standards of living. In that case, we distribute income to the entire resident population: we refer to this convention as “per capita” wealth or income.

In general, average income per adult follows a similar path to average income per capita, but there are exceptions. In countries that have recently undergone a severe demographic transition (e.g., Mexico), the trends are markedly different. Income per capita is a better measure of how much individuals can consume, whereas income per adult is a better measure of how much they earn. Our primary objective is to measure the latter, but both concepts are complementary.

1.1.2 Individualistic Adults

The “individualistic” series distribute income or wealth to each person individually according to individual earning and ownership. When doing so, it makes a lot more sense to restrict ourselves to the adult population, since most children earn zero income. We refer to such series as “individualistic adults” series.

In the “individualistic adults” series, observed labor income and pension income is attributed to each individual recipient. This is easy to do in individual taxation countries, like today’s United Kingdom, since 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, we only observe the total labor or pension income reported by both spouses. This is especially true

[2] To extent that some children earn income, they tend to do so on behalf of their parent or guardian.

of countries with joint tax filing that only provide tax tabulations to researchers: see chapters 5 and 7 for how to make proper estimations in such cases.

Attribution to individuals is 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 common assets and own assets. Therefore, 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.[3]

Self-employment income presents a particular dilemma, as its remuneration includes both a return to labor and a return to capital, which are by construction difficult to disentangle. (For this reason it is referred to as mixed income.) To be consistent with the calculations above, we allocate to each spouse 50% of the estimated capital share of mixed (self-employment) income, whereas we allocate 100% of the estimated labor share of mixed income to the self-employed adult individual. In all of these cases, whenever better data sources are available, we try to offer a more sophisticated treatment of the distribution of capital income within households.

Similar problems arise for taxes and transfers. In countries with joint tax filing, we cannot easily separate which spouse pays the income tax. Even in individual tax filing countries, some benefits may be means-tested at the household level, and may depend on the household's number of children, etc. These cannot be easily separated, either. Therefore, we disaggregate these flows equally between household members. In the end, our "individualistic adult" series may include a fair amount of income that is in fact shared among household members. One way to avoid this problem is to focus on incomes that are easier to individualize, such as labor and pension income. We at least try to provide "individualistic adult" series for these types of income.

1.1.3 Equal-split Adults

The "equal-split adults" series distribute income to all adult individuals, while splitting income equally within a couple or a household. This is often less demanding in terms of data, especially in countries with joint tax filling.

[3]Note that we also split 50–50 the capital income of couples with "civil union contracts" — such as the Pacs in France — who according to French law file joint returns and report a single capital income amount (just as for married couples).

However, equal-split series raise an important secondary question, namely 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., grandparents living with their adult children), this can make a significant difference, especially at the bottom of the distribution (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. In terms of data availability, narrow equal-split series are easier to compute when using tax microdata as the primary source. Broad equal-split, on the other hand, is easier to calculate from surveys, which generally identify households but rarely identify couples within households. The household is also a concept that is clearly defined within the SNA, while the couple is a more elusive notion: in countries with joint taxation, people may face different tax incentives to declare themselves single or in a couple.

For now, DINA series have relied on narrow equal-split measures in countries that primarily rely on tax microdata (France, United States), and broad equal-split in countries that rely more heavily on surveys (China). This should be kept in mind when making comparisons between countries.[4] Moving forward, whenever the data allow, researchers should try to provide both series.[5] In particular, researchers working on countries with tax microdata, who also in general have access to survey microdata, should make the effort to import the household structure from the survey data (i.e., tax units per household, along the income gradient) into their tax data, to calculate broad equal-split series. The opposite — calculating narrow equal-split in countries with survey data only — is usually more difficult.

1.1.4 Per Capita Equal-split

The "per capita equal-split" series distribute income equally to all household members, including children.[6] As discussed in section 1.1.1, this convention is a better measure of how much people can consume, as opposed to how much they earn. This constitutes useful, complementary information to our equal-split adult benchmark. This convention is also the one used by the World Bank for its PovcalNet database. While the WID has until now focused on equal-split adult series, we intend to progressively integrate per capita equal-split series in the future.

[4] See, for example, the discussion in Piketty, Yang, and Zucman (2019) and the comparison between DINA series for China, France and the United States.

[5] See Blanchet, Chancel, and Gethin (2022) for an example.

[6] By construction, this requires using the "broad" family unit discussed in section 1.1.3.

1.1.5 Equivalence Scales

The literature on inequality has often used various "equivalence scales" to account for household structure in inequality estimates. The motivation behind equivalence scales is that households with many members face economies of scale, so that the resources needed (e.g., for heating, size of dwelling, transportation, etc.) by a household with four members are less than four times the resources needed by a single-individual household.

Equivalence scales define a number of "consumption units" within households. Each household member then gets an income equal to the household income divided by the number of consumption units. There are several equivalence scales that have been suggested in the literature on inequality. The two most common ones being the square root equivalence scale and the modified OECD equivalence scale. In all equivalence scales, the number of consumption units grows more slowly than the size of the household, so that people who all earn the same income are better off in a three-person household than a two-person household. The modified OECD equivalence scale also takes age into account, so that, for example, a single parent with a child appears about 15% better off than a childless couple for the same household income (and the same number of individuals).

The DINA project does not use equivalence scales for two reasons. The first one is practical. Equivalence scales introduces nonlinearities that make it harder to connect aggregate levels of income and wealth with their distribution. If we attribute to each individual their equivalized income, then the sum of all incomes across all individuals no longer sums up to aggregate income. Similarly, average equivalized income across all individuals is neither equal to aggregate income divided by the number of people, nor is it equal to aggregate income divided by the number of consumption units. Equivalence scales therefore prevent us from cleanly breaking down national income and its growth across population groups. In our view, it negatively impacts the readability and ease of use of indicators, and somewhat defeats the purpose of connecting micro and macroeconomic estimates in the first place.

The second reason is conceptual. The primary aim of DINA is to measure income, wealth and their distribution. In contrast, equivalence scales are meant to approximate an individual "welfare" concept. While this distribution of "welfare" is related to income and wealth, it also incorporates many other dimensions that are outside of the project's scope. Indeed, the DINA project does not explicitly attempt to measure how well the income of individuals suits their needs — which depend not only on their household's size, but also on their location, their tastes, their health status, etc., all of which are partly endogenous to distribution of income and wealth themselves. To properly account for all these factors would be an extremely ambitious project — one that equivalence scales only partially address — and the resulting computations would likely be fragile.

1.2 Prices and Currency Conversion

To compare levels of income and wealth, it is necessary to correct for differences in the price level. The price level evolves over time (because of inflation or deflation) and differs between countries (because of the use of different currencies, or because similar goods and services are sold at different prices in different places). Accounting for these effects raises numerous challenges.

A large number of indicators provided by the WID, such as the share of income and wealth received by various groups, or wealth-to-income ratios, are not sensitive to the overall price level. For other indicators, the database provides market exchange rates, PPPs, and a price index. They are used to perform the aggregations of income and wealth at the regional and at the world level in the database, and to assist users in their own computations. There is one variable for the price index, which we call the national income price index. For market exchange rates and PPPs, we give conversion factors from local currencies to US dollars, euros, and yuan. Hence, there are three variables for market exchange rates. Similarly, there are three variables for PPPs.

By default, all series in the WID are expressed at constant prices in the LCU. Users can then use the price index, the exchange rates and the PPPs to convert to current prices or to different currencies. In practice, there are many conceptual challenges involved in these computations: this section emphasizes what these are, and justifies our methodological choices.

1.2.1 Price Index

The basic identity to move from constant prices to current prices is the following:

$$\text{current prices} = \text{constant prices} \times \text{price index}$$

Several concepts of the "price index" can be used in the formula above. The two main ones are the GDP deflator and CPI.[7] We use the GDP deflator as our preferred price index, following

[7] Other indexes are sometimes computed by national statistical institutes or central banks. For example, in the United States, the Bureau of Economic Analysis computes the PCE, a deflator that relates solely to the consumption part of GDP. The Federal Reserve uses the core PCE, which further excludes commodities with volatile prices such

Piketty and Zucman (2014), although we resort to the CPI when no other data is available. We call it the national income price index, and use it to deflate all the monetary series.

The differences between the CPI and the deflator are twofold. First, there is their definition. The deflator measures the price level of domestic production (as defined by the GDP), while the CPI measures the price level faced by consumers. Therefore, the deflator only considers domestic goods and services, and nothing that is imported, while the CPI does include foreign goods. But the deflator measures the price level of all goods and services, while the CPI only considers what is bought by consumers.

The second difference has to do with methodology. There are three potential biases involved in comparing prices over time: how to account for new goods (the "new good bias"), how to account for quality improvement (the "quality bias"), and how to account for changes in the consumer's basket of goods and services when the prices change (the "substitution bias"). Statisticians have made progress in addressing the new good bias and the quality bias for both the CPI and the deflator. But for the substitution bias, more progress has been made on the deflator, essentially by switching to the chain-linking method (Lequiller and Blades, 2014; Piketty and Zucman, 2014). The CPI, on the other hand, is still generally calculated as a simple Laspeyres index.[8]

This is why we use the deflator as our preferred measure of the price level, especially when looking at distant periods of time. There can be important short-run differences in the inflation rate as measured by the two indexes, mostly because the CPI incorporates volatile foreign goods such as oil. More importantly, those differences can accumulate over time so that, in the United States, the total inflation since the 1960s has been significantly lower according to the deflator than the CPI. That is because the deflator is better at taking the substitution bias into account. That being said, comparing income or wealth levels between time periods that are centuries apart is fraught with conceptual difficulties, and it could be argued that such comparisons do not always make sense. This is why we often emphasize ratios or shares over income levels. While we provide some price indexes over very long time periods for convenience, users should be

as food and energy, as it prefers to focus on measures of long-run inflation. However, these indexes do not have internationally agreed definitions and they are not calculated in many countries. They serve some very specific purposes such as fine-tuning monetary policy. Here we focus on the two most common indexes, for which we have data in a sufficiently large number of countries.

[8] See Piketty and Zucman (2014, data appendix, p. 38), for an in-depth discussion of the differences centered on the United States, and Lequiller and Blades (2014) for a detailed description of chain-linking methods.

aware of these issues.[9,10]

1.2.2 Currency Conversion

The WID provides currency conversion factors to convert from one currency to another, in the same way that the national income price index can be used to convert from one year to another. For each country, we give a conversion factor in the form "local currency unit per foreign currency," where the foreign currency can be the US dollar, the euro or the yuan. Because series are in local currency by default, users can convert them using:

$$\text{series in foreign currency} = \frac{\text{series in local currency}}{\text{currency conversion factor}}$$

The conversion factor is here to reflect the relative purchasing power of each currency. We provide two conversion factors: market exchange rates and PPPs.

1.2.2.1 Market Exchange Rates and Purchasing Power Parities

The most obvious currency conversion factor is the market exchange rate. For example, if we can buy 8 yuan with 1 euro, then the euro/yuan exchange rate is 8, meaning that one euro offers eight times the purchasing power of one yuan. Therefore, to convert a series from yuan to euros, we divide it by 8.

The problem with market exchange rates is that they only reflect the relative purchasing powers of currencies in terms of *tradeable* goods: 1 euro or 8 yuan may get you the same amount of crude oil on international markets, but it doesn't necessarily let you buy the same amount of

[9] Also note that choosing only one price index is a simplification. We might wish to use a different price index for each series, one that would reflect the composition of goods and services of the series we are trying to deflate. That is, in fact, what some national accounts do: they compute GDP and each of its component both in prices and in volume. The deflator is the ratio the two. That ratio for a subcomponent of GDP (e.g., consumption) will not be equal to that of the whole GDP. The issue with that approach is that it ends up changing the relative shares of GDP components. Quantities that are dimensionless (e.g., saving rates) will be different if we compute them from volume or price quantities. The issue becomes critical with chain-linking, because it breaks the additivity of national accounts, so that standard accounting equations no longer hold exactly. The discrepancies are small and usually have no economic meaning, but they make it harder to get a consistent global view of the economy, and also to apply economic models that rely on these identities. Bringing inequality into the mix raises even more questions (should the income of the top 1% be deflated differently from the rest?) All of this conflicts with the WID's ultimate goal of providing a consistent view of macro aggregates and their individual distribution. Our solution of using only one price index has the advantage of simplicity and transparency, while yielding results extremely close to more sophisticated approaches.

[10] Technically, our preferred prince index should be deflator of national income rather GDP, given our focus on the former. But this deflator is rarely available, and in practice the differences would be minimal.

food in Paris or Beijing. Yet the latter is more relevant to people's standards of livings. Because non-tradeable goods and services tend to be cheaper in emerging economies (a phenomenon known as the Penn or Balassa–Samuelson effects) market exchange rates underestimate their standard of living.

The solution to that problem is to use PPPs. PPPs are conversion factors that estimate more accurately the actual relative purchasing power of currencies. It is now standard practice to use PPPs rather market exchange rates as the standard currency conversion factor for comparing income levels between countries. Yet the use of market exchange rates can still be justified in some contexts (see chapter 8 for a discussion).

In the WID, we use PPPs by default to convert all series. We also provide market exchange rates (period average), and users can request them if they want to. We provide two versions of regional estimates: one at market exchange rates and one at PPPs.

By definition, two countries sharing the same currency will have the same market exchange rate[11]. They will not, in general, have the same PPP. For example, all countries of the euro area have a market exchange with the euro equal to one by definition. But their PPP conversion factors can be sensibly different. Indeed, the European countries have different relative price levels, so that one euro doesn't get you the same amount of goods and services in Paris, Berlin or Vilnius, even though you can pay with the same currency in all three cities. The PPP conversion factors take this into account. Therefore, using PPPs for euro area countries leads to monetary values that are expressed in a "Europe-wide euro," which reflects the average purchasing power of the euro in all of Europe, and is different from the "French euro," the "German euro," or the "Lithuanian euro."

1.2.2.2 Estimation of Currency Conversion Factors

Market exchange rates are readily available as quoted values on the currency market. We only need one value per year, so we take end-of-year values. For older periods, before the existence publicly available forex data, we use data from the World Bank that goes back to 1960.[12]

[11] This is not true for Eurozone countries prior of them adopting the Euro. In this cases, we rely on UN SNA exchange rates of Euro to USD -which are available on the country level basis from 1970 in some cases-. To compute a unique pre-1999 Euro to USD exchange rate that can be used for any country, we do so averaging the exchange rate of Framce, Italy, Germany, Spain and The Netherlands, weighted by their GDP. An identical procedure is used for PPP.

[12] In some countries (e.g., Venezuela in recent years), the official market exchange rate has been maintained artificially low by local authorities. This official rate does not correspond to any economic reality, and leads to an unrealistically high GDP per capita is US dollars. Therefore, in these cases, we assume that the actual market exchange rate with the US dollar has followed the relative evolution of price indexes compared to the United States,

For PPPs, we use estimates from the OECD for OECD countries, and from the World Bank otherwise. Compared to exchange rates, PPPs raise more methodological difficulties. In many non-OECD countries, their computation started fairly recently. Moreover, new PPP data arrives at infrequent intervals: at the time of writing, the last three rounds of the ICP — which is in charge of their computation — for which data were released happened in 2005, 2011 and 2017.

How, then, can we know the proper PPP conversion factors for years that fall outside of an ICP round? The solution is to extrapolate them based on the relative evolution of the price index (PI) in both countries (McCarthy, 2013). For example, we estimate the PPP in 2023 from the PPP in 2017 using:

$$\text{PPP}_{2023}^{\text{home/foreign}} = \text{PPP}_{2017}^{\text{home/foreign}} \frac{\text{PI}_{2023}^{\text{home}}/\text{PI}_{2017}^{\text{home}}}{\text{PI}_{2023}^{\text{foreign}}/\text{PI}_{2017}^{\text{foreign}}} \tag{1.1}$$

When we have access to several rounds of the ICP for a given country, two solutions are possible. The first one is to use solely the most recent round and discard previous ones: this the solution adopted by the World Bank. The other approach is to adapt formula (1.1) to interpolate between two ICP round: this is the solution adopted by the Penn World Tables (Feenstra, Inklaar, and Timmer, 2015). We follow the World Bank's approach. Indeed, the estimation of PPPs is difficult, and some changes from one round to the next are purely methodological. For example, the 2005 round stirred controversy because it led to a significant downward revision of China's and India's GDP (see Maddison, 2010). The methodological issue was solved in the 2011 round. Using both the 2005 and 2011 rounds of the ICP to compute India's and China's PPP would therefore lead to an overestimation of growth over the period for spurious methodological reasons. Moreover, given that the estimation of PPPs started fairly recently in most countries, we would have to extrapolate them for most of the series anyway. But the World Bank's approach has its flaws, too: in particular it forces us to revise former PPPs values every time a new ICP round is released, which retroactively changes values of, say, the world GDP. We still favor it because it seems more important to us to preserve the real growth rates of individual countries.

1.2.2.3 Combining Currency Conversion and Price Indexes

We use the price index to compare income levels between different periods, while we use currency conversion factors to compare them between countries. But how can we compare two incomes from two different places and two different years? The issue arises for example if we want to plot national income series of two countries on the same graph.

---

as in formula (1.1).

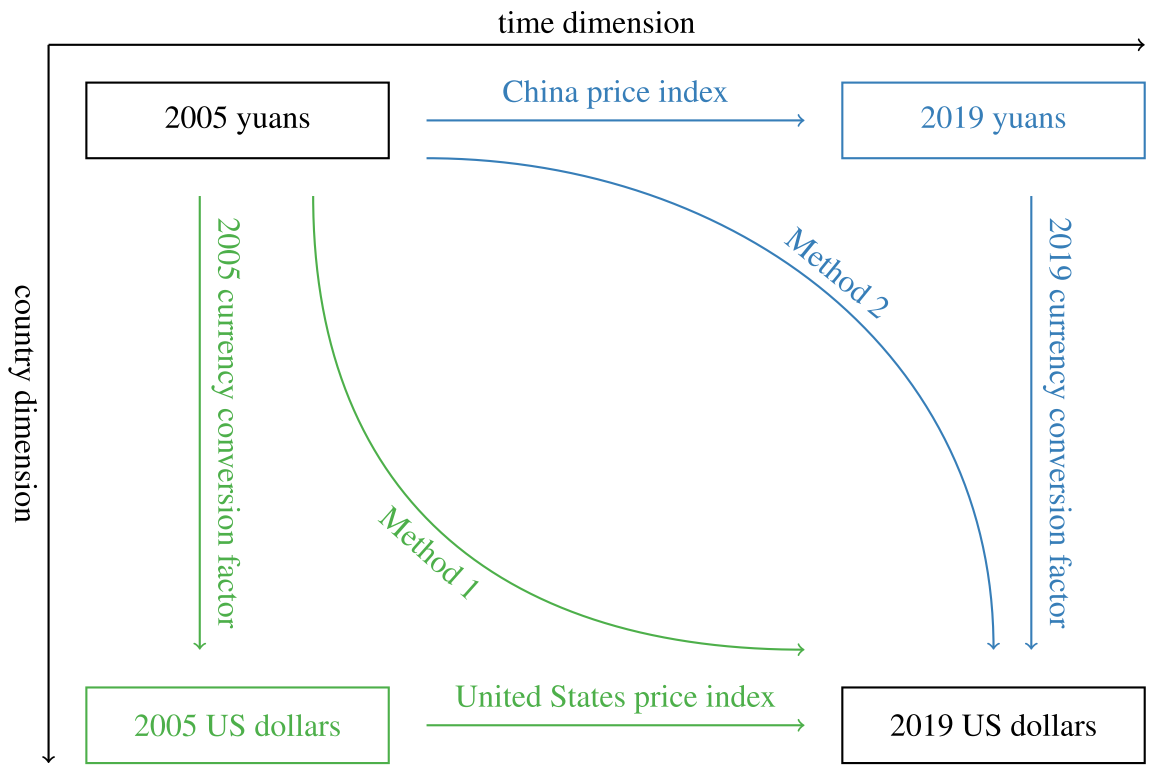


Illustration of the two different ways of converting monetary series both over time and between countries. We have to move across two dimensions: time and country. Method 1 (green) moves across countries first and across time second. Method 2 (blue) does the opposite.

Figure 1.1: Two Methods for Converting Series Across Time and Currencies

There are two ways to proceed, as illustrated by figure 1.1, and both methods may give different results. As an example, assume that we want to convert from 2005 yuan to 2019 US dollars.

With the first method (green), we first move along the country dimension, so we apply the 2005 currency conversion factor to get an amount in 2005 US dollars. Then we apply the United States price index to convert the 2005 US dollars in 2019 US dollars. The second method (blue) starts with the time dimension. So we first use the Chinese price index to convert in 2019 yuan, and then the 2019 currency conversion factor to convert into 2019 US dollars.

Which method is the most appropriate? When using the PPP conversion factors, both are strictly equivalent. This is because the extrapolation procedure in formula (1.1) makes both approaches identical by construction. Recall that, by default, WID series are in LCU. Therefore, in the WID, we use the PPP conversion factor of the reference year to convert the entire series (following method 2 in figure 1.1).

But when using the market exchange rates, the methods are not identical anymore. The right approach depends on the motivation for using market exchange rates over PPPs. The website uses method 2, both because of slightly better data availability, and lower volatility for series in individual countries.[13]

1.3 Measures of the Distribution

The primary objective of the WID is to provide the best and most comprehensive estimates of the distribution of income of wealth possible. Given that objective, we try not to focus on any specific indicator of inequality, but instead to describe distributions of income and wealth with as much detail as possible.

In our view, to focus on a single indicator — for example the well-known Gini coefficient — is not sufficient to provide an adequate picture of inequality. We stress that the problem is not specifically about the Gini, or about any other indicator. The problem is that inequality cannot be reduced to a single number.

Imagine, for example, that a country simultaneously sees a reduction in poverty, a rise of top incomes, and a shrinking of the middle class. These evolutions have contradictory effects on most synthetic indicators of inequality. For a given indicator, they might cancel out, giving the impression that the underlying distribution has not changed, when in fact it did. Indicators that implicitly give a lot of weight to the bottom of the distribution will show a decrease in inequality, while those that put more weight on the top will show an increase. But all of them miss the real underlying story because the answer to the question "has inequality increased?" does not have an unequivocal answer. There are, in effect, three stories happening at the same time. The goal should be to study them separately, not to average them out.

This is why we prefer to describe distributions in their entirety, and then let users use whatever level of detail and whatever indicator suits their needs. In terms of presenting the results, we favor showing the income shares of three main groups (the bottom 50%, the middle 40% and the top 10%). These three groups map relatively well to the idea of a lower, middle and upper class. They summarize changes happening to the overall distribution (such as the ones from the example above) fairly well. In practice, synthetic indicators can be approximated quite precisely by a weighted average of these three shares. We also emphasize the very top groups (top 1%, top 0.1%, etc.) since they can represent a macroeconomically significant share of income and

[13]When converting macroeconomic series with a varying market exchange rate over time, the evolution of the exchange rate tends to overdetermine the evolution of the series in the short run.

wealth. We provide Gini coefficients in the database as a convenience but encourage users to look further.

But once again, the full data included in the database goes beyond these key groups and describes the entire distributions with as much detail as possible. There are two broad types of output for DINA series: synthetic microfiles and series by g-percentiles.

1.3.1 Synthetic Microfiles

Synthetic microfiles are files containing microdata that are representative of the distribution of income and wealth for the entire population. They are organized similarly to survey or tax microdata: each row is an observation that represents individuals from the full population, while columns correspond to variables (such as components of income and wealth). Each observation is anonymized, and in practice combines data from a variety of sources (hence the term "synthetic"). All income components are made consistent with macro aggregates.

Synthetic microfiles reproduce the joint distribution of income and wealth in the true population, and can be used like any other microdata. They provide the greatest flexibility to the final user. They can easily be used to perform arbitrary changes in the population unit, cross-tabulations, demographic breakdowns, etc. In addition, micro-files can be used as a resource for further analysis by various actors from research institutions and civil society — for instance, to simulate tax reforms.

While microfiles are our preferred output for inequality data, there are several obstacles to their production. The greatest is confidentiality. Much of the data — both survey and administrative — used to estimate distributions is subject to strict confidentiality agreements, preventing the diffusion of the final files. In some countries, the microdata is not even made available to the researcher, and thus the underlying methodology for producing the series does not lend itself to the production of microfiles. For now, the microfiles have been publicly released for the United States.

1.3.2 Series by g-percentiles

Even when the production of microfiles is not possible, we give as much information as possible by providing series that describe the thresholds, averages and shares for every "generalized percentiles." There are 127 of these generalized percentiles (or g-percentiles):

- 99 percentiles from $p = 0\%$ to $p = 99\%$,

- 9 tenths of a percentile from $p = 99\%$ to $p = 99.9\%$,
- 9 hundredths of a percentile from $p = 99.9\%$ to $p = 99.99\%$,
- 10 thousandths of a percentile from $p = 99.99\%$ to $p = 100\%$.

These indicators are sufficient for most users. In particular they can be used to compute any inequality indicator with a great level of precision, locate oneself or others in the distribution, and perform arbitrary aggregations of countries.

1.4 Regional aggregation

While most of our data are provided at the country level, we also produce regional inequality series that follow a consistent aggregation scheme (Table 1.1). These regional estimates are designed to be as comparable and homogeneous as possible across space and time. The geographical structure builds on 57 territorial units (48 main countries and 9 residual regions) selected to balance global coverage with the historical continuity of statistical sources. This framework allows us to construct long-run measures of inequality. The aggregation provides a coherent basis for studying regional patterns of inequality and their evolution in a global perspective. Note that the 48 main countries make about 85-90% of the world GDP (both in market exchange rate and purchasing power parity) and population throughout the 1800-2025 period (see Nievas and Piketty, 2025b, Figure 1).

Table 1.1: Main geographical aggregation used across the database

Region (N)	**Countries / regions**
East Asia (5)	China, Japan, South Korea, Taiwan, Other EASA
Europe (11)	Britain, Denmark, France, Germany, Italy, Netherlands, Norway, Spain, Sweden, Other W. EUR, Other E. EUR
Latin America (6)	Argentina, Brazil, Chile, Colombia, Mexico, Other LATAM
Middle East / North Africa (8)	Algeria, Egypt, Iran, Morocco, Saudi Arabia, Turkey, UAE, Other MENA
North America / Oceania (5)	USA, Canada, Australia, New Zealand, Other NAOC
Russia / Central Asia (2)	Russia, Other RUCA
South / South-East Asia (9)	Bangladesh, India, Indonesia, Myanmar, Pakistan, Philippines, Thailand, Vietnam, Other SSEA
Sub-Saharan Africa (11)	DR Congo, Ethiopia, Kenya, Ivory Coast, Mali, Niger, Nigeria, Rwanda, Sudan, South Africa, Other SSAF

Notes. This aggregation of 57 core territories (48 main countries + 9 residual regions) is used consistently across the database and underpins the construction of regional estimates. Residual "Other" units pool countries with limited or discontinuous historical coverage to preserve comparability and homogeneity over time.

The regional framework introduced in this version represents an updated definition of the one used in previous releases (Table 1.2). This revision expands the geographical coverage of earlier versions, moving from 33 to 57 core territories. While country-level data remain the foundation of our work, this new aggregation broadens the reach of our regional estimates across space and time. These core territories are consistent with the results presented in the World Inequality Reports and in the Global Justice Project.

Compared with the previous structure, the new definition broadens coverage across nearly all regions. East Asia now includes South Korea and Taiwan alongside China and Japan; Europe extends to Northern and Eastern European countries such as Denmark, the Netherlands, and Norway; and South and South-East Asia now encompasses a wider range of countries, including Bangladesh, Myanmar, Pakistan, Philippines, Thailand, and Vietnam. Similar extensions were made in the Middle East and North Africa, and in Sub-Saharan Africa, where the group now covers major economies such as Nigeria, Kenya, and Ethiopia. These additions provide a more balanced representation of global diversity while maintaining the long-run comparability of earlier series.

Table 1.2: New vs. Old Definition of Core Territories Used in the Database

Core Territories: Old Definition (33 core territories = 24 main countries + 9 residual regions)		**Core Territories: New Definition** (57 core territories = 48 main countries + 9 residual regions)	
East Asia (3)	China, Japan; Other EASA	**East Asia (5)**	China, Japan, South Korea, Taiwan; Other EASA (OA)
Europe (8)	Britain, France, Germany, Italy, Spain; Sweden, Other W. EUR; Other E. EUR	**Europe (11)**	Britain, Denmark, France, Germany, Italy, Netherlands, Norway, Spain, Sweden; Other W. EUR (OC), Other E. EUR (QM)
Latin America (6)	Argentina, Brazil, Chile, Colombia; Mexico, Other LATAM	**Latin America (6)**	Argentina, Brazil, Chile, Colombia, Mexico; Other LATAM (OD)
Middle East / North Africa (4)	Algeria, Egypt, Turkey; Other MENA	**Middle East / North Africa (8)**	Algeria, Egypt, Iran, Morocco, Saudi Arabia, Turkey, UAE; Other MENA (OE)
North America / Oceania (5)	USA, Canada, Australia, New Zealand; Other NAOC	**North America / Oceania (5)**	USA, Canada, Australia, New Zealand; Other NAOC (OH)
Russia / Central Asia (2)	Russia; Other RUCA	**Russia / Central Asia (2)**	Russia; Other RUCA (OC)
South / South-East Asia (3)	India, Indonesia; Other SSEA	South / South-East Asia (9)	Bangladesh, India, Indonesia, Myanmar, Pakistan, Philippines, Thailand, Vietnam; Other SSEA (OI)
Sub-Saharan Africa (2)	South Africa; Other SSAF	**Sub-Saharan Africa (11)**	DR Congo, Ethiopia, Kenya, Ivory Coast, Mali, Niger, Nigeria, Rwanda, Sudan, South Africa; Other SSAF (OJ)

Notes. The new definition expands the earlier 33 core territories to 57 units. Codes denote residual regions standardized with the new definition (OA, OC, OD, OE, OH, OI, OJ, QM).

Key Points

- DINA series use consistent units of observation over time and between countries. Our benchmark series use the "equal-split adult." They distribute income or wealth to the entirety of the resident adult population (by which we mean 20 and older) of an economic territory. Income is split equally between members of a couple. When the data allows it, we also consider alternative units, such as "individualistic adults" (income not split between household members) and "per capita equal-split" (income distributed to children as well as adults).
- We do not use "equivalence scales" for both conceptual and practical reasons. In particular, they would prevent us from cleanly linking macroeconomic aggregates to the individual distribution of income or wealth.
- The GDP deflator is our benchmark price index for all DINA series. Over long periods of time, comparing quantities that are independent of the price level (wealth/income ratios, share) can be more informative.
- We provide PPPs and market exchange rates to convert between currencies. In our view, they both provide complementary information. We extrapolate PPPs over time based on the most recent ICP round based of the relative evolution of prices between countries.
- Monetary series in the WID are always given in the local currency at the latest year's prices. Therefore, to convert them using PPPs or market exchange rates, we only use the most the recent PPP or market exchange rate, and apply it to the whole series.
- The database does not focus on any specific indicator, but instead seeks to provide a complete description of the distribution from the bottom to the very top.

Chapter 2

Income Concepts

Beyond the challenge of the observation unit discussed in chapter 1, another limitation of inequality series based only on raw tax data, such as those that used to be published in the WTID, is the lack of homogeneity of income concepts. In the DINA project, we overcome this issue by defining harmonized income definitions for both pretax and post-tax income inequality. This chapter explains and discusses these concepts.

Most WTID series were constructed using a so-called "fiscal income" notion, meaning the total income that is or should be reported on income tax declarations (before any specific deduction allowed by fiscal legislation).[1] This concept naturally varies with the tax system and the legislation of the country and year under consideration. It does not correct for the fact that some forms of income (in particular some forms of capital income) are legally not subject to personal income taxes, and do not appear on income tax declarations.[2] As a consequence, the "fiscal income" concept can have series breaks and comparability problems.

In contrast, the income concepts we use in DINA series are defined in the same manner in all countries and all time periods, regardless of the fiscal legislation of any given country and year. As we will explain below, the four basic pretax and post-tax income concepts by which we

[1] "Fiscal income" is broader and slightly more homogeneous 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 a ceiling). In the case of France, like in most countries, the raw tax data generally use the concept of "taxable income" (post-deduction income), and a number of corrections were applied so that WTID series refer to "fiscal income" (pre-deduction income). Although the "fiscal income" concept in WTID series is broader than "taxable income," it is not sufficiently broad and homogeneous over time and across countries.

[2] Sometimes, 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.

measure income inequality are anchored to the comprehensive notion of "national income" — and are designed to fit within the macroeconomic national accounting framework that calculates GDP, set forth in the 2008 SNA standard.[3]

Our decision to use national accounts income (and wealth) concepts for distributional analysis certainly does not mean that we believe that these concepts are perfectly satisfactory or appropriate. We are aware that official national accounts have many shortcomings and need to be improved. In fact, we view the DINA project as a contribution to that improvement. This is why we occasionally correct SNA concepts when we consider them inadequate for our purpose (for example with respect to the measurement of foreign income, see below), and in some cases make suggestions for changes to future SNA definitions. But whenever the concepts from the WID depart from the SNA, we make it explicit and justify our choices.

The main 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, and which (at least in principle) can be applied to all countries and years for which we have data, independently of country-specific and time-specific legislation and data sources. These concepts do need to be refined, but to propose amendments and improvements, the best is to start from them, use them and modify them when needed. The alternative would be to start from scratch and propose entirely new definitions of income, output and wealth, which is neither realistic nor desirable.

Moreover, by using national accounts concepts and producing distributional series based on them, we hope we can help address what we view as their main blind spot: namely, that they do not provide any information about the extent to which the different social groups benefit from growth. To close the gap between inequality measurement and growth measurement in national accounts could also help close the gap between popular perceptions of economic growth (the actual distribution of growth among a population) and its macroeconomic measurement in the headline (quarterly and annual) GDP growth statistics.

[3]In some countries, or for some earlier years, available national accounts series still follow the earlier system of international guidelines, namely SNA 1993 (or the European version, the 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. The main innovation between SNA 1993/ESA 1995 and the 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").

2.1 Aggregate Income

This section describes the aggregate income flows to be distributed to individuals. We do not yet explain how these aggregates should be distributed in pretax or post-tax income: this is left for section 2.2.

We illustrate this section with data from the "Sequence of accounts" tables provided by the SNA 2008 guidelines (United Nations, 2009). The actual amounts reported in this "Sequence of accounts" table progression do not refer to any real setting, but the overall structure is meant to be broadly representative of the national accounts of advanced economies, and to demonstrate how the components of these accounts must necessarily fit together.[4]

While we try to give a description of SNA concepts that is sufficient to understand and create DINA series, a complete overview of SNA concepts is beyond the scope of this document. For that, we refer the reader the complete guidelines (United Nations, 2009), or to Lequiller and Blades (2014) for a more pedagogical treatment of the topic.

2.1.1 Net National Income

The aggregate income concept of reference in the WID is NNI (B5n, S1).[5] This marks a subtle but important departure from many macroeconomic databases that tend to focus on GDP (B1g, S1).

B1g, S1	**Gross Domestic Product (GDP)**	**1854**
	Plus: Net Foreign Income (NFI)	15
	Of which: officially recorded	10
D1, S2	*Of which: net compensation of employees*	4
D4, S2	*Of which: net property income*	6
D3–D2, S2	*Of which: subsidies less taxes on production and imports*	0
	Of which: income from offshore tax havens (estimated)	3
	Of which: reinvested earnings on foreign portfolio investment (estimated)	2
P51c, S1	Minus: Consumption of Fixed Capital (CFC)	222
B5n, S1	**Equals: Net National Income (NNI)**	**1647**

Adapted from the SNA 2008 "Sequence of accounts" (United Nations, 2009). See online appendix for the construction of the table.

Table 2.1: GDP and Net National Income

[4]When we depart from the SNA definition, we similarly choose arbitrary numbers for the new or modified components when needed, while making sure that the orders of magnitude remain plausible.

[5]The codes in parentheses refer to the official SNA codes for the transaction and the sector. Whenever we mention an SNA item, we provide these codes to help connect the discussion to official sources.

Table 2.1 presents the movement from GDP to net national income. In this example, GDP, the standard measure of total annual production in a nation's economy, is equal to 1854. From this amount, we remove 222 of CFC: the income that is estimated to be lost due to the depreciation of the capital stock. Then we add 15 of NFI: this corresponds to the income generated abroad but which accrues to domestic residents, minus the income generated in the domestic economy but which accrues to foreigners. In the present case, the former (inflow) is larger than the latter (outflow), so net foreign income is greater than zero. We then arrive at net national income, equal to 1647.

While the GDP statistic is more frequently discussed by academics and the general public, we believe NNI is a more meaningful statistic for our purposes: indeed, CFC is not earned by anyone, whereas inflows and outflows of foreign income do affect the amount of money that residents get. While GDP and NNI usually follow each other, there are notable examples of countries and eras where the two diverge. GDP is particularly sensitive to assumptions (and legal labels) on the location of production — a notion that can become murky in an era of globalization and internationally integrated production chains. In countries such as Ireland or Luxembourg, GDP growth in recent years has been coupled with large outflows of capital income, a phenomenon that can be at least partly attributed to tax avoidance by multinational corporations.[6] As an indicator of income rather than production, national income is not sensitive to such issues.[7]

2.1.1.1 Net Foreign Income

We notably depart from official statistics in the measure of net foreign income. In the example above, officially recorded net foreign income is equal to 10. To that, we add two components. The first, "income from offshore tax havens," is an estimate of income flows that go unrecorded in official statistics because they come from assets sheltered in tax havens and not reported to any authorities. Here we have a value of 3.

Our second foreign income component which transcends the official statistics is "reinvested earnings on foreign portfolio investment" (2 in our example). This item represents a modification of the way the SNA records foreign income flows. In broad strokes, the income that a company retains after having paid its suppliers, its employees, its shareholders, and its corporate income

[6] For example, Ireland officially estimated its real GDP growth in 2015 to be +26%. This number stirred controversy, as it is believed to be the sole result of a few large multinational corporations relocating their intangible assets in Ireland for tax purposes.

[7] Net foreign incomes compensate any change in GDP caused by different assumptions about the localization of production.

tax bill is what we call "undistributed profits" or "retained earnings." This flow is part of national income. However, imagine that a company in country A has some undistributed profits, but is actually owned by residents of country B. How are they recorded in national accounts? Under the current SNA guidelines, it depends. If they belong to an investment fund, then these profits are implicitly distributed to their owners, be they in country A or B: this transaction is recorded in item D443. If the ownership takes the form of FDI, meaning that the residents of country B have some direct control over the company, then the undistributed profits of the company are attributed to country B: this transaction is recorded in item D43. However, if the ownership takes the form of portfolio investment, meaning that the residents of country B do not have a direct control over the company's decisions, then the SNA currently considers that the entire flow of undistributed profits belongs to the national income of country A, not country B.

One rationale behind the SNA's differential treatment of portfolio and direct investment is that portfolio investors have no direct control over the company's undistributed profits, and therefore this flow of income does not "belong" to them. In our view, this rationale is problematic for at least three reasons. The first one is that the frontier between portfolio and foreign direct investment is somewhat arbitrary. The current guidelines use sharp cutoffs: ownership of more than 10% is categorized as direct investment, while ownership below 10% is categorized as portfolio. This distinction can be useful for certain purposes, but obviously there is little practical difference between 9.9% and 10.1% ownership. The second reason is that, while portfolio investors do not exert direct control, they can still decide whether or not to invest in a company. If they choose to invest in a company that keeps its profits as retained earnings rather than distributing them as dividends, it must be because they see an interest in doing so. The third reason, closely tied to the second, is that undistributed profits can be construed as latent capital gains. If a company owned by portfolio investors uses its retained earnings to accumulate assets, it mechanically increases the value of the company, which constitutes an income (in the Hicksian sense) to the investors because it increases their wealth. This income accruing to foreign residents (of country B) does not allow any domestic resident (of country A) to either consume or save.[8]

The concern used to be second order, because the rise of foreign portfolio investment is a relatively recent phenomenon. Until the early 1980s, more than 90% of United States foreign equity assets were FDI investments, so that almost all foreign profits were *de facto* included in national income. But since then, the share of FDI has gradually dropped and is currently close

[8]Note that net asset positions, as recorded by the IMF, do not treat foreign direct investment differently from portfolio investment, so that in practice foreign flows get recorded somewhat differently from the stocks.

to 50%.

Our correction estimates both the flow of foreign retained earnings that accrue to residents and the flow of domestic retained earnings that accrue to foreigners. The difference between these two items leads to our adjustment. In the future, we would recommend that SNA guidelines move toward a more homogeneous treatment of foreign income flows, as we do here.

Note that the flow of net foreign retained earnings that we estimate is after payment of the corporate tax, even though is it considered a "primary" income. This is in line with the current treatment of reinvested earnings on foreign portfolio investment. Yet in fact, we believe that the SNA guidelines should move even further, and also include corporate income taxes in cross-border primary income flows. In the SNA, there is currently no cross-border flows of corporate income tax payments. Direct investment income is after-tax. Portfolio investment income is after tax as well (and after retained earnings). But in reality, to the extent that they are foreign-owned, some of the corporate taxes paid by domestic corporations are paid by foreigners. Conversely, domestic residents pay foreign corporate taxes on their share of foreign corporate profits. This is true both for direct equity investments and portfolio equity investments. Conceptually, the corporate tax should always be treated as undistributed corporate profits. If some of these profits accrue to foreigners, then some corporate taxes should also be assigned to foreigners. If we include undistributed corporate profits in portfolio investments, then we should also include corporate taxes in portfolio investments. This would make it possible to study important questions, such as the extent to which some countries are able (or not) to make non-residents pay corporate income taxes (e.g., resource-rich countries).

We explain in chapter 4 how we make both of these corrections. The remainder of foreign income is sourced directly from the official statistics. The two main components are the compensation of employees (D1, S2, paid minus received) and property income (D4, S2, paid minus received). "Compensation of employees" accounts for the wages and salaries of domestic residents working abroad, minus the wages and salaries of foreigners working in the domestic economy. "Property income" accounts for the incomes generated by foreign assets that residents own, minus the incomes generated by domestic assets that foreigners own.[9] Finally, there is the foreign balance of "subsidies less taxes on production and imports" (D3–D2, S2, paid minus received), which is zero in this example, and usually small.

[9]Note that the paid/received convention has to be interpreted from the perspective of the rest of the world, so that inward income flows are "paid" and outward ones are "received."

2.1.1.2 Consumption of Fixed Capital

The other step in moving from GDP to NNI is to subtract the CFC (P51c, S1). It is an estimate of the reduction of the value of fixed assets used in production, resulting from physical deterioration, normal obsolescence or normal accidental damage.[10]

It is important to deduct depreciation from our measure of income because the consumption of fixed capital does not allow anyone to either consume or accumulate wealth, and therefore does not constitute income. To include it would artificially inflate the income of capital owners. That being said, we are aware that the measurement of the consumption of fixed capital is imperfect and raises many issues. Unlike most SNA components, it cannot be directly observed, and has to be estimated. Its estimation is complex and not necessarily homogeneous between countries. One of its most important limitations is that it does not usually include the consumption of natural resources. In other words, CFC tends to overestimate both the levels and the growth rates of national income, which in some cases should be much lower than those obtained for GDP. In the future, we plan to gradually introduce such adjustments to the aggregate national income series provided in the WID. This may introduce significant changes both at the aggregate and distributional level.

Another issue with CFC statistics is that, unlike GDP or net foreign income, many countries do not estimate the consumption of fixed capital at all, which complicates the study of net national income across the world. We include estimates of it in the WID when the information is missing from official sources (see section 4.1.2.2 for the methodology), so that we can at least provide coherent estimates of net national income in all countries. Those estimates should be viewed as provisional and subject to revision.

2.1.1.3 Global consistency of income flows: correcting for measurement errors and hidden wealth

One well-documented anomaly in balance of payment statistics is that when summing up net foreign capital income at the global level, the result tends to consistently be negative rather than zero. This implies that the world as a whole is paying more than what is receiving from foreign investments, which is impossible. The explanation offered in the literature is that negative imbalances are primarily caused by accounting errors and assets hidden in offshore tax havens, which are recorded as liabilities but never as assets (Zucman, 2013, section 3.3.3), and the

[10] The consumption of fixed capital is conceptually similar to the idea of depreciation, if slightly distinct in national accounting from corporate accounting (see Lequiller and Blades, 2014).

income accrued from them.

In the WID, proportional adjustments are made to ensure that net foreign capital income, net foreign wealth and rest of items sum up to precisely zero at the global level, as it should naturally be, which is conditional on a comprehensive accounting of all economies. The correction simply adjusts credits and debits to the average of total global credits and total global debits. For a detailed discussion of this method and its comparison with other alternatives, see Nievas and Piketty (2024).

As shown in Nievas and Piketty (2024), the results that we obtain by using this simple proportional adjustment method are very close to those that would be obtained by using direct estimates of the share of assets hidden in tax havens by every country (see e.g. Alstadsæter, Johannesen, and Zucman (2018). The advantage of the proportional adjustement method is that it can be applied in a consistent manner to subcategories of assets and income flows, which is not the case with direct estimates of tax haven assets (which are not available for all subcategories).

This adjustment allows us to identify systematic winners and losers across countries even in the absence of bilateral data. Building on it, research finds that the excess yield - i.e. the gap between returns on foreign assets and returns on foreign liabilities - has increased significantly for the top 20% richest countries (population weighted) since 2000. This translates in net income transfers from the poorest to the richest equivalent to 1% of the GDP of top 20% countries (and almost 2% of GDP for top 10% countries), alleviating the current account balance of the latter while deteriorating that of the bottom 80% by about 2-3% of their GDP (Nievas and Sodano, 2024).

2.1.2 Sectoral Decomposition of Net National Income

In the SNA, national income can be decomposed as the sum of the income of the different sectors of the total economy: households, corporations, and the government. The complete decomposition for our example is shown in table 2.2.

This decomposition is a measure of the proportions of national income that accrue to the different sectors of the economy, before the operation of direct taxes and transfers (in the SNA conception of “direct” taxes, more on this below). The household sector (S14) captures the majority of that income. It is often pooled together with a much smaller sector, non-profit institutions serving households (S15). Several countries do not attempt to separate these two sectors. The corporate sector (S11+S12) can be split between financial (S12) and non-financial corporations (S11), although this distinction is not of primary importance for our purposes. Finally, the general

B5n, S14 + S15	Plus: net primary income of households and NPISH	1362
B5n, S14	*Plus: net primary income of households*	*1361*
B5n, S15	*Plus: net primary income of NPISH*	*1*
B5n, S11 + S12	Plus: net primary income of corporations	114
B5n, S11	*Plus: net primary income of non-financial corporations*	*99*
B5n, S12	*Plus: net primary income of financial corporations*	*15*
B5n, S13	Plus: net primary income of the general government	171
B5n, S1	**Equals: net national income**	**1647**

Adapted from the SNA 2008 "Sequence of accounts" (United Nations, 2009). See online appendix for the construction of the table.

Table 2.2: Sectoral Decomposition of Net National Income

government sector (S13) captures the remaining income. The primary income of each of these sectors can be broken down further, and the following sections will describe in greater detail the elements of this decomposition that are useful to the production of DINA series.

The WID contains the breakdown of national income at the finest level possible for all the countries for which data is available through standard portals (UNdata, OECD.Stat, etc.) or estimated by researchers. We systematically combine all the data at our disposal while ensuring the overall consistency of the database (see chapter 4 for methodology). That being said, official national accounts are fairly rudimentary in many developing countries and some developed countries. Sometimes the accounts do not include the level of detail that we need. In countries where national accounts are too fragile and where other data sources allow us to estimate income and wealth series that are more adequate and consistent, we recommend using these complementary data sources, and we update our series accordingly. Again, we do not pretend that the concepts and estimates we provide are perfectly satisfactory, but we try to make the best of the available data in a clear and transparent way.

Note that in the decomposition of table 2.2, net foreign income is already distributed to the different sectors, which is why it is not listed in this decomposition (which sums to national income). Given that we have estimated additional components of national income ourselves, we must allocate them to the various sectors for the decomposition and accounting identities to remain valid. We attribute missing income from tax havens to the household sector, and we attribute reinvested earnings on foreign portfolio investment to non-financial corporations.

2.1.2.1 Income of Households and NPISH

2.1.2.1.1 Income of Households

The net primary income of households (B5n, S14) is made up of four main components: compensation of employees (D1), (distributed) property income (D4), net operating surplus (B2n), and net mixed income (B3n). Table 2.3 shows the details of this decomposition in our example.

Uses			Resources		
			B2n	**Operating surplus, net**	**69**
			B3n	**Mixed income, net**	**53**
D4	Property income	41	D1	Compensation of employees	1154
D41	*Interest*	*14*	*D11*	*Wages and salaries*	*954*
D45	*Rents*	*27*	*D12*	*Employers' social contributions*	*200*
			D4	Property income	126
			D41	*Interest*	*14*
			D42	*Distributed income of corporations*	*20*
			D43	*Reinvested earnings on foreign direct investment*	*3*
			D44	*Other investment income*	*30*
			D441	*Investment income attributable to insurance policyholders*	*20*
			D442	*Investment income payable on pension entitlements*	*8*
			D443	*Investment income attributable to collective investment fund share holders*	*2*
			D45	*Rents*	*21*
				Income from offshore tax havens	*3*
B5n	**Primary income, net**	**1361**			

Adapted from the SNA 2008 "Sequence of accounts" (United Nations, 2009). See online appendix for the construction of the table.

Table 2.3: Allocation of Primary Income, Households

The largest item is the compensation of employees (D1). Note that it represents the total cost of labor, including employers' social contribution, both real and imputed (see secondary distribution of income below for details). This constitutes an important discrepancy with many sources that include only employee (and only "actual") social contributions.

Property income (D4) appears both in the "uses" and the "resources" columns. On the resources side, interests (D41) includes the interest received by households on their financial investments. The distributed income of corporations (D42) corresponds to dividends (D421) and withdrawals from income of quasi-corporations (D422). This latter component corresponds to income distributed to households from "quasi-corporations," i.e., entities that are not corporations in the legal sense (and therefore do not distribute dividends in the legal sense either) but which have strong similarities to them and which the SNA therefore classifies in the corporate sector. Therefore, there can be a conceptual difference between dividends in the national accounts and dividends in other sources (e.g., in tax data). In general, item D422 is small, but there are exceptions, since the treatment of quasi-corporations is not universally consistent. In Italy, for example, any unincorporated enterprise with more than five employees is considered a quasi-corporation (Lequiller and Blades, 2014, p. 179). Therefore, item D422 includes a lot of income that would be considered mixed income (B3n, see below) in other countries. To the extent that this discrepancy is meaningful, researchers should take it into account and, if need be, provide a more consistent treatment of quasi-corporations. In the future, we would encourage countries to harmonize their definition of the corporate/household sector split. Other investment incomes (D44) contains capital income flows that are attributed to households even though they are not able to consume them freely. This includes income from life insurance policies (D441), pension funds (D442), and investment fund shares (D443). Note that item D442 includes investment income "earned" by employers' unfunded pension plans, which in our view should rather be reclassified as corporate savings. We recommend making this adjustment when possible (see section 3.1.1 for an extended discussion). Finally, rents (D45) includes income receivable by owners of land and natural resources (e.g., subsoil assets). Importantly, this does not include payments received from leased dwellings, vehicles, or other types of fixed capital: those are included in operating surplus and net mixed income (B2n+B3n, see below). We also add a new item, income from offshore tax havens, which captures the income estimated to be received by households who hide assets in tax havens (see section 2.1.1.1).

On the "uses" side, we have some capital incomes that must be subtracted from households' primary income. Interests (D41) are the interest payments made by households when they take out consumer or housing loans, as well as the interest payments made by unincorporated enterprises on their borrowing. Rents (D45) are the payments made by unincorporated enterprises when they use land and natural resources.

The last two items to highlight are net operating surplus (B2n) and net mixed income (B3n). Many countries pursue slightly different approaches to distinguish these entities, or only report both items combined. In the SNA guidelines, the operating surplus of households includes both actual rental income received by landlords and imputed rental income of owner-occupied housing. In practice, countries are not fully consistent in their reporting. Germany, for example, only includes imputed rents in operating surplus (B2n), and puts actual rental income in mixed income (B3n). Researchers should address each situation on a case-by-case basis.[11] In the future, we hope that countries will harmonize their definitions, or better yet provide an explicit breakdown of actual and imputed rents to avoid any confusion. Net mixed income (B3n) corresponds to the remuneration of the owners of unincorporated enterprises. This remuneration comprises both wages and profits (labor and capital), and it is not possible to distinguish these two components.

Table 2.4 details selected items from the secondary distribution of income account of the household sector. This account records the transactions that relate to the redistribution of income, be it through social insurance or social assistance. We focus on three main items and their subcomponents: current taxes on income wealth, etc. (D5), social contributions (D61), and social benefits (D62). Note that in the SNA, there is an additional item named "other current transfers" (D7). This item covers an heterogeneous set of transactions. It makes sense to include some of them (for example remittances) in our distributional series, but not others (for example parking fines). And in many cases (insurance premiums, charitable giving), the issue is debatable. Given the wide variety of cases, we prefer to deal with the issue on an case-by-case basis. For now, by default, we exclude "other current transfers" from our concepts — so we *de facto* treat them as a form of consumption — but we may occasionally decide to include some of it in income when it is genuinely important (e.g., remittances in some countries).[12] In the future, we intend to provide better treatment of the issue. In general, this item is relatively small (especially in net terms) so its impact should be limited. But this should be kept in mind when comparing some of our series with other sources: to avoid confusions with the SNA concept of "disposable income" (B6n, which includes D7), we use the term "secondary income" for the balancing item at the end of table 2.4.

The item "current taxes on income, wealth, etc." (D5) groups direct taxes such as the personal income tax (from all levels of government) or wealth taxes (when they exist). Note that in many cases, certain taxes that could arguably be considered direct taxes — as in the classifications of public revenues of the OECD or the IMF — are actually classified in the SNA as other taxes on production (D29). For example, property taxes on businesses are classified under D29, while a residence tax on households is still classified under D5. Hence, the business property tax

[11] See country-specific studies and national methodology publications of the statistical offices. In some cases, supply and use tables provide an explicit decomposition of imputed and actual rents, which was used by Blanchet, Chancel, and Gethin (2022) to harmonize definitions in Europe.

[12] Note that according to the official definition of remittances from the IMF, they also include item D1 (compensation of employees) of foreign income, which is included in our concepts.

Uses			Resources		
			B5n	**Primary income, net**	**1361**
D5	Current taxes on income, wealth, etc.	178	D62	Social benefits other than social transfers in kind	384
D61	Net social contributions	333	*D621*	*Social security benefits in cash*	53
D611	*Employers' actual social contributions*	181	*D6211*	*Social security pension benefits*	45
D6111	*Employers' actual pension contributions*	168	*D6212*	*Social security non-pension benefits in cash*	8
D6112	*Employers' actual non-pension contributions*	13	*D622*	*Other social insurance benefits*	279
D612	*Employers' imputed social contributions*	19	*D6221*	*Other social insurance pension benefits*	250
D6121	*Employers' imputed pension contributions*	18	*D6222*	*Other social insurance non-pension benefits*	29
D6122	*Employers' imputed non-pension contributions*	1	*D623*	*Social assistance benefits in cash*	52
D613	*Households' actual social contributions*	129			
D6131	*Households' actual pension contributions*	115			
D6132	*Households' actual non-pension contributions*	14			
D614	*Households' social contribution supplements*	10			
D6141	*Households' pension contribution supplements*	8			
D6142	*Households' non-pension contribution supplements*	2			
B6n+D7	**Secondary income, net**	**1234**			

Adapted from the SNA 2008 "Sequence of accounts" (United Nations, 2009). See online appendix for the construction of the table.

Table 2.4: Secondary Distribution of Income, Households

is included in the primary income of the general government (see section 2.1.2.3), while the household residence tax is included in the primary income of the household sector and then transferred to the government in the secondary distribution of income account. In our view, this distinction is not necessarily meaningful, and in some cases, we recategorize some (conceivably "direct") taxes from D29 to D5. The item D5 also excludes inheritance taxes, which are classified under D9 as a capital transfer but rarely disaggregated from other (non-tax) types of capital transfers.

Social contributions (D61) in the "uses" column includes households' payments to social insurance schemes. It is divided in several subcomponents, indicating decompositions across several dimensions. First, there is the distinction between employers' and households' social contributions. The former are paid by employers on behalf of employees, while the latter are paid directly by employees.

Second, we can distinguish actual from imputed employers' contributions. Actual social contributions are observable payments made to social security: they usually represent the majority of social contributions. Nonetheless, sometimes employers provide social security benefits directly to their employees, without going through a social security fund. In these cases, national accountants estimate the social contributions that employees would have to pay to receive these benefits, and attribute them to households. This is true, in particular, of many civil servant pensions, which are directly paid by the general government. These imputed contributions are typically absent from other sources (e.g., tax data) and must therefore be imputed in those sources as well.

Third, households' social contributions are divided into "actual" and "supplementary." The former refers to actual payments (like for employers' contributions). The latter refers to a payment that was initially included as part of item D442 ("investment income payable on pension entitlements") in the primary allocation of income account. The property income earned by pension funds was indeed distributed to households as part of their primary income, but since this income is used to pay out benefits to households, it has to be removed when looking at the redistribution of income, which is done here.

Fourth, every contribution can be separated between "pension" and "non-pension" components. Indeed, while pensions often form the bulk of social insurance systems, social contributions also pay for benefits related to health, family, etc.

Social benefits other than social transfers in kind (D62) includes all direct (cash) transfers to households. Social security benefits (D621) and other social insurance benefits (D622) refer to all social insurance (i.e., employment-related) social benefits. They are further broken down into "pension" and "non-pension." The other (non-employment-related) cash transfers are categorized under social assistance benefits in cash (D623). Note that some national statistical offices consider certain benefits like housing benefits to be "in kind" rather than "in cash" and therefore classify them as individual consumption expenditure (P31, S13) from the government sector. In the context of DINA series, we view it as more meaningful to consider them as if they were cash transfers, and recategorize them as such.

In practice, the level of decomposition of table 2.4 for social contributions and social benefits is rarely available from the main sources. When a certain level of decomposition is not available from national accounts, researchers can resort to other sources. For developed countries, the

OECD Social Expenditure and Tax Databases provide breakdowns along similar lines. For many European countries, the Eurostat provides similarly detailed information.

2.1.2.1.2 Income of NPISH

The primary income of NPISH (B5n, S15) is always a very small fraction of national income. In fact, many countries do not report this sector separately, but pool it with the household sector (S14+S15). This aggregation is based on the view that since NPISH are financed by households and provide services to them, their account can essentially be assimilated into those of households. When this is the case, net operating surplus (B2n) in the sector includes not only incomes from leased dwellings and imputed rents of owner-occupiers as before, but also the operating surplus of NPISH.

Given the usual size of the NPISH sector, this usually makes little difference. Conceptually, in the context of DINA, it nonetheless makes more sense to treat the income of NPISH as a separate component whenever possible.

2.1.2.2 Income of Corporations

Table 2.5 decomposes the primary income of corporations, both non-financial and financial (B5n, S11+S12), which is essentially two components: net operating surplus (B2n) and property income (D4).

Uses			Resources		
			B2n	**Operating surplus, net**	**169**
D4	Property income	302	D4	Property income	245
	Reinvested earnings on foreign portfolio investment	3		Reinvested earnings on foreign portfolio investment	5
B5n	**Primary income, net**	**114**			

Adapted from the SNA 2008 "Sequence of accounts" (United Nations, 2009). See online appendix for the construction of the table.

Table 2.5: Allocation of Primary Income, Corporations

Net operating surplus (B2n) measures the profit of corporations. On top of this, corporations pay and receive property income from other corporations and other sectors. In general, this yields a negative net property income (D4) because corporations pay more property income to other sectors than they receive. In this case it amounts to −57. In the WID, we also include the estimated item "reinvested earnings on foreign portfolio investment" (see section 2.1.1.1), which we add to the primary income of non-financial corporations.

Uses			Resources		
			B5n	**Primary income, net**	**114**
D5	Current taxes on income, wealth, etc.	34	D61	Social contributions	279
D62	Social benefits other than social transfers in kind	267			
B8n+D7	**Corporate savings, net**	**92**			

Adapted from the SNA 2008 "Sequence of accounts" (United Nations, 2009). See online appendix for the construction of the table.

Table 2.6: Secondary Distribution of Income, Corporations

Table 2.6 shows the secondary distribution of income account. Current taxes and income, wealth, etc. (D5, S11+S12) correspond to the corporate tax. The two other items, social contributions (D61) and social benefits other than social transfers in kind (D62) have been discussed at length in the section 2.1.2.1.1 on the household sector.

The presence of these items in the accounts of the corporate sector might come as a surprise. Indeed, in general, social contributions and social benefits are paid to and received from the general government sector. However, some large firms sometimes set up their own social insurance plans: the corresponding contributions and benefits appear here. Note that several countries do not follow the SNA in this regard. In Australia, for example, imputed employer contributions are recorded within the compensation of employees (D1) but are not shown as being paid back to employers (D61, S11+S12). Researchers should attempt to correct such discrepancies whenever they arise.

The balancing item (i.e., the residual that offsets "uses" and "resources") is net corporate savings (B8n+D7, S11+S12). In short, it measures retained earnings after taxes and transfers, while the primary income measures retained earnings before taxes and transfers.[13]

2.1.2.3 Income of the General Government

Table 2.7 decomposes the primary income of the general government (B5n, S13). It is composed of three main components: net operating surplus (B2n), taxes less subsidies on production and imports (D2–D3) and property income (D4).

The net operating surplus (B2n) of the general government is zero in this example. This is in fact a convention of national accounting. Because the output of the government sector is not sold at meaningful market prices, it is valued at cost (i.e., compensation of employees for labor,

[13] As explained before (see section 2.1.2.1.1), we do not explicitly account for "other current transfers" (D7).

Uses			Resources		
			B2n	**Operating surplus, net**	**0**
D3	Subsidies on production and imports	44	D2	Taxes on production and imports	235
D4	Property income	42	D4	Property income	22
B5n	**Primary income, net**	**171**			

Adapted from the SNA 2008 "Sequence of accounts" (United Nations, 2009). See online appendix for the construction of the table.

Table 2.7: Allocation of Primary Income, General Government

and consumption of fixed capital for capital). As a result, its net operating surplus is zero by construction. In practice, however, a small part of the government sector does operate as market enterprises (for example, some water-supply units) and report some profits or losses that go into the net operating surplus. But this amount always remains small.

Taxes less subsidies on production and import (D2–D3) represent the bulk of the government's primary income. It represents "indirect" taxes that are collected during the production process. As result, the value added that pays for them is never distributed to any factor of production. Direct taxes, on the other hand, are paid out of the remuneration of a factor of production. In practice, some taxes (such as property taxes) are classified as indirect taxes even though they could easily be construed as direct taxes (see discussion in section 2.1.2.1.1). Nowadays, in most countries, sales and VATs represent the majority of indirect taxes, while tariffs were more important in the past. It is important to account for these taxes, since they are part of national income. They often represent an important fraction of government revenues, and their incidence can have significant distributional consequences.

Finally, property income (D4) is the difference between the income that the government receives on its assets (when they are used in a market context) and the interest that the government pays on its debt. This item is usually negative.

Table 2.8 describes the secondary distribution of income account. Among these different items, the current taxes on income, wealth, etc. (D5), social contributions (D61) and social benefits other than social transfers in kind (D62) that we observe here, are the counterpart in the government account of their values in the household and the corporate sector accounts, and have been discussed above in sections 2.1.2.1.1 and 2.1.2.2.

That leaves the item final consumption expenditure (P3), which represents direct government spending on behalf of individuals. It is split between individual consumption expenditure (P31) and collective consumption expenditure (P32). The former corresponds to spending that can

Uses			Resources		
			B5n	**Primary income, net**	**171**
D62	Social benefits other than social transfers in kind	112	D5	Current taxes on income, wealth, etc.	213
P3	Final consumption expenditures	352	D61	Social contributions	50
P31	*Individual consumption expenditures*	*184*			
P32	*Collective consumption expenditures*	*168*			
B8n+D7	**Surplus/deficit, net**	**-30**			

Adapted from the SNA 2008 "Sequence of accounts" (United Nations, 2009). See online appendix for the construction of the table.

Table 2.8: Secondary Distribution of Income, General Government

unmistakably be attributed to specific individuals (e.g., health, education), the latter to spending that benefits the collectivity at large (e.g., defense, police, general administration). For some types of spending (e.g., culture) the frontier is somewhat arbitrary.[14]

The difference between government income (primary income, current taxes, social contributions) and spending (transfers in cash and final consumption) is what we call the surplus/deficit of the government, and is included in national income. Note that this item differs from what is usually used to measure government deficits, i.e., net savings (B8n) or net lending/borrowing (B9n), due to the exclusion of other current transfers and capital transfers from the analysis (see section 2.1.2.1.1).

2.1.3 Factor Shares and the Sectoral Decomposition of GDP

The WID decomposes income between labor and capital. To be precise, we decompose national income into four components: pure labor income, pure capital income, mixed income, and taxes on products. Pure labor income includes the compensation of employees of households. Pure capital income includes the property income and operating surplus (of households and NPISH), the undistributed profits of corporations, and the property income of the government. That last term is usually negative because it is mostly composed of interest payments on the government debt. Its inclusion means that interest income paid by the government is not included in the capital share. (Other conventions are possible, see Piketty and Zucman (2014) for a discussion.)

[14] The SNA 2008 introduced the concept of adjusted disposable income that classifies individual consumption expenditure (P31) as a transfer to households (D63, equal to P31). This changes the disposable income of the household sector, but given our focus on national income, is just a categorization with no concrete consequence in terms of broad aggregates or distribution.

Mixed income is the income of the self-employed, which combines the remuneration of labor and capital. The rest corresponds to the indirect taxes, which are paid to the government before they can explicitly accrue to any given factor of production.

Assuming a fixed proportion of mixed-income to be attributed to capital and labor, we define the capital share in the total economy as:

$$\text{Total capital share} = \frac{\text{Gross operating surplus of all sectors} + 40\% \text{ of gross mixed income}}{\text{GDP} - \text{taxes on products}}$$

By definition:

$$\text{Labor share} = 1 - \text{capital share}$$

We apply a simple 60–40 split between labor and capital for gross mixed income. This convention serves as our benchmark for the sake of transparency and comparability, but we do not claim it is the only or even the best approach. In practice, the appropriate definition of labor and capital shares depends on context, and there is no single correct answer. For this reason, we also provide the full sectoral decomposition, allowing users of the WID to apply alternative conventions if they wish.

Net-of-depreciation series are generally preferable, since replacement costs reduce the income accruing to capital owners. However, estimates of depreciation rates vary across countries and are not always comparable. For this reason, we present both gross and net series. Similarly, we provide both domestic measures (excluding net foreign income) and national measures (including net foreign income), as both perspectives are relevant depending on the analytical focus. Because consumption of fixed capital is deducted from capital income, gross capital shares are always larger than that their net counterparts. Note that taxes on products are always excluded from the denominator, so that all shares are expressed relative to factor prices rather than market prices; we do not attribute taxes on products to either capital or labor.

In order to provide meaningful global comparisons of labour and capital shares, we stress the need to study factor shares together with the structure of production by institutional sectors. In particular, the share of macroeconomic income originating from the household sector (including self-employed individuals) rather than from the corporate or government sector varies enormously across countries and between the poorest and the richest regions. To do so, we decompose GDP into three institutional sectors and production taxes: the corporate sector (including financial and non-financial corporations: S11 and S12), the government sector (S13) and the household sector (including households and non-profit institutions serving households: S14 and S15). The value added by each sector can itself be divided into the compensation of employees, gross operating surplus, consumption of fixed capital (CFC) and net operating surplus, as well as mixed income from self-employment in the case of the household sector.

The value-added of the three institutional sectors sums up to GDP at factor-price, adding net production taxes results in GDP at market prices.

$$\text{Corporate capital share} = \frac{\text{Corporate gross operatig surplus}}{\text{GVA of the corporate sector}}$$

Based on the sectoral decomposition we define the capital (labor) share within the corporate sector as the share of operating surplus (wages) of total factor price value-added of the corporate sector. The value added of the corporate sector is the sum of wages paid by the corporations and their operating surplus.

For more details on the computation of factor shares and their sectoral decomposition please refer to Dietrich et al. (2025).

2.2 Income Distribution

We aim to provide income distribution estimates for at least four broad concepts of income: pretax factor income, pretax post-replacement income (which we will generally abbreviate as pretax income), post-tax disposable income, and post-tax national income.

For pretax series, the key difference between pretax national income and pretax factor income is the treatment of pensions (and other insurance-based social benefits), which are counted on a distribution basis for pretax national income and on a contribution basis for pretax factor income (more on this in sections 2.2.2 and 2.2.3). We tend to favor the "pretax national income" as our benchmark concept to measure inequality before redistribution. But we stress that the "pretax factor income" inequality series also provides useful and complementary information. Our series are constructed so that both pretax national income and aggregate pretax factor income match the net national income aggregate.

Among post-tax series, our "post-tax disposable income" series aims to describe post-tax, post-transfer inequality for the population's actual perceived budget constraints, while excluding in-kind transfers such as health and education and other public spending (as these may impact purchasing power and disposable income only indirectly). For this reason, aggregate post-tax disposable income can be substantially less than aggregate national income — typically around 70% of national income, since in-kind transfers and public spending typically represent about 30% of national income. Our "post-tax national income" series include all in-kind transfers and public spending (using various procedures to attribute them to individuals, see section 2.2.5) so that aggregate post-tax national income is equal to aggregate national income.

It is worth stressing that these four concepts are all defined in terms of income concepts from the 2008 SNA, and are measured using data that is, to the extent possible, consistent with the amounts reported in the national accounts.

2.2.1 Inequality Measurement and the National Accounts

Let us start by explaining some of the main novelties introduced in DINA series, compared to most other standard inequality datasets. Our focus on national income as a whole implies that we must find and distribute income components (in particular those outside of the household sector) that are excluded from traditional sources, be they survey data or tax data. The three most important ones are the "undistributed profits" (or "retained earnings") of corporations, the indirect taxes, and the government's final consumption expenditures.

2.2.1.1 Undistributed Profits and the Corporate Tax

The income concept of undistributed profits, or retained earnings, refers to income remaining in the corporate sector after the payment of wages, interest, dividends, and, for post-tax concepts, the corporate income tax. These profits are a component of national income and GDP as defined by the SNA. And yet, most types of inequality data (including raw personal income tax data series) do not include them. This has created several problems for studying inequality.

First, it means that significant parts of national income, including parts of the tax base of governments, have been excluded from inequality statistics. In particular, it makes it difficult (if not impossible) to meaningfully analyze the redistributive impact of corporate income taxes, which are one of the most important taxes for the very wealthy.

Second, if excluding retained earnings, series regularly show sudden (artificial) changes in inequality, because the frontier between the household and the corporate sector is porous. Whether income stays within a company or gets redistributed to shareholders can often depend on tax incentives, and a delay in shareholder payouts (even or especially if the delay is indefinite) is not a real decrease in the standard of livings of the individuals concerned.

A good example of this case is the Tax Reform Act of 1986 in the United States. Companies in the United States can be registered as *S-corporations* or as *C-corporations*. Large companies are usually C-corporations. This means that they are subject to the corporate tax, and can pay dividends to their shareholders, who are then subject to the income tax. Smaller companies tend to be S-corporations, which are not subject to corporate taxes and do not pay dividends. Instead, the firm's profit is directly included in the income tax base of their owner, who pays the federal income tax on that profit directly. For companies at the margin, whether to be a C-corporation or an S-corporation is essentially a tax arbitrage decision. With the 1986 Tax Reform Act, the top marginal personal income tax rate became lower than the corporate income tax rate. As a result, many companies had a strong incentive to switch their legal form from a C-corporation to an S-corporation, so that profits would flow to individuals directly rather than face the corporate income tax. In the next two years, a large amount of capital income entered the federal (personal) income tax base at the top of the personal income distribution, resulting in a large, sudden increase on inequality. Yet this series break was the result of a legal change with no real economic significance.

Other countries have had similar experiences. A similar reform in Norway was analyzed by Alstadsæter, Jacob, et al. (2017), taking advantage of the extremely detailed nature of Norwegian data. They show that around the reform, we observe large series break in inequality and income mobility as measured using the income tax. Yet once we attribute firms' profits to their owners, these effects disappear. While the data is more limited in most countries, DINA series attempt to make similar corrections to get rid of such artificial effects. By including retained earnings in our measure of personal income, we avoid any sudden jumps in income from the artificial exclusion of real (if not yet "realized") earnings.

Third, to exclude retained earnings would mean excluding flows that do constitute income in the Hicksian sense, because they increase household wealth. Indeed, when a firm accumulates cash, its value mechanically increases. This increase in value constitutes a latent capital gains for its owners. Raw tax-based series would sometimes address this problem by including realized capital gains. However, realized capital gains are a poor measure because the realization of capital gains is itself the results of various tax incentives (see Robbins (2018) for an extended discussion). Moreover, in some cases, capital gains are never taxed (e.g., the "stepped-up basis" loophole in the United States tax code). While we do not include all capital gains because they are not part of national income (and their inclusion would introduce a lot of volatility, also see Robbins (2018)), by including retained earnings we take into account the regular structural factors behind irregularly realized capital gains.

Note that sometimes governments own a significant part of the corporate sector. In these cases, it is important to make sure that only part of retained earnings is distributed to individuals. Indeed, the SNA classifies publicly owned companies in the corporate sector as long as they provide goods or services at "economically significant prices." Therefore, the retained earnings of publicly owned companies are part of the primary income of the corporate sector, not the government sector. There are at least two ways to distinguish between income of publicly owned and private companies: when financial balance sheets are available, we can look at the fraction of equities owned by the household sector among the equities owned by both the household and the general government sector, and only distribute the corresponding fraction to individuals. Sometimes, countries explicitly distinguish in their national accounts publicly owned from privately owned corporations within the corporate sector. In this case only retained earnings from the latter should be included. In both cases, the fraction of retained earnings not attributable to households should be treated like a property income to the government (see section 2.2.2).

Note that the question of retained earnings attributable to foreigners (and foreign retained earnings attributable to residents) has been resolved thanks to our inclusion of the additional item "reinvested earnings on foreign portfolio investment" in foreign income (see section 2.1.1.1). For countries with an extremely unbalanced net foreign income position (see section 4.1.2.3), researchers should pay special attention to this adjustment, as it can have a significant impact on inequality.

2.2.1.2 Taxes on Production

Taxes (less subsidies) on production (D2–D3) constitute the majority of the primary income of the government. One reason for including them in pretax income is that the frontier between production taxes and direct income and wealth taxes (D5) is somewhat arbitrary — so that it is unclear why we should deduct the former and not the latter. Thus, 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, be they production taxes or direct taxes. Production taxes also constitute an important source of revenue for governments: excluding them from analysis would bias the comparison of tax levels and progressivity among countries with different tax systems.

Nowadays, in most countries, sales taxes and VATs constitute the majority of production taxes. (Taxes on international trade were a more important indirect tax in earlier eras.) Note that, following the convention of national accounts, they must be added to observed income levels in order to reach a consistent pretax income concept, rather than subtracted from it.

This convention is somewhat at odds with intuition and with the practice of certain microsimulation studies. However, in practice, this is the only way of providing a treatment that is consistent with direct taxes (which are included within household primary income), and which avoids double-counting. Indeed, in the SNA, taxes on products are already deducted before the value added is used to remunerate factors of production (unlike direct taxes). If we were to remove taxes on products from household income, we would effectively be removing them twice. This is why we choose to distribute them as part of pretax income. There are several ways of doing so, and we discuss our preferred approach in section 2.2.2.

2.2.1.3 In-Kind Transfers and Collective Expenditures

A third issue in distributing all of national income is in distributing government final consumption expenditure in post-tax income. In-kind transfers (P31) and collective expenditures (P32) are the two components of government consumption (P3). In-kind transfers typically include health and education spending, while collective expenditure includes public spending such as national defense and street lighting (or, 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 2.2.5), which is why we also report results for "post-tax disposable income," which excludes these components.

2.2.2 Pretax Factor Income

Pretax factor income, which for simplicity we often refer to as "factor income," is equal to the sum of all pretax income flows accruing directly or indirectly to the owners of the factors of production (labor and capital), before the operation of the tax and transfer system (including indirect taxes), and before the operation of social insurance systems.

One good aspect of the factor income concept is that it is relatively easy to compute using national accounts data, and it is reasonably homogeneous across countries. Its main drawback, however, is that old-age individuals generally have little factor income, so that cross-sectional inequality of factor incomes look artificially large in countries and time periods with an older population. One way to overcome that issue, when the data allows it, is to restrict the analysis to the working-age population.

Aggregate pretax factor income is by construction equal to net national income, and can be defined as the sum of the primary income of the household sector (B5n, S14), the primary income of the corporate sector (B5n, S11+S12, i.e., pretax undistributed profits), the primary income of the government sector (B5n, S13), and the primary income of NPISH (B5n, S15).

The primary income of the household sector (B5n, S14) corresponds to the full compensation of employees (including employers' social contributions, both actual and imputed), self-employment income, and distributed capital income (including imputed rents). The primary income of the corporate sector (B5n, S11+S12) corresponds to pretax undistributed profits and includes the corporate income tax. The primary income of the government (B5n, S13) is made up of two main components: taxes on production (D2–D3, S13) and property income (D4, S13).

We naturally distribute the compensation of employees and property income to the respective owners of these household income flows. We distribute the income of NPISH — usually a very small amount — proportionally to the rest of factor income. The remaining questions concern undistributed profits, production taxes and the property income of the government.

2.2.2.1 Distribution of Undistributed Profits

The undistributed profits are recorded in national accounts as (1) primary income of corporations (B5n, S11+S12) and (2) dividends and withdrawals from quasi-corporations (D422, S14), if any. We recommend to distribute these undistributed profits in proportion to stock ownership, be they held directly or indirectly, in privately or publicly traded companies. More refined imputations are possible whenever more detailed data is available — for example, data that matches the income tax records of individual shareholders to the firms they own (see also section 5.3.2).

2.2.2.2 Distribution of Taxes

The general principle that guides the allocation of taxes in pretax DINA series is that taxes are paid by the factor of production on which the tax depends, in line with distributional tax analysis (Saez and Zucman, 2019). We stress that this approach is different from nominal incidence: for example, employers' (actual and imputed) social contributions are allocated to employee compensation even though they are nominally paid by employers, because the amount of the contribution depends on the employee's wage. Our approach is also different from a complete tax incidence study, one that would try to establish a full counterfactual of what the economy would look like if the taxes were changed or did not exist. Such studies can be useful, but they are not the point of DINA — or of national accounts in general. First, they require many behavioral assumptions and the effects are far from obvious. (Indeed, for every type of tax, there is a contentious literature on the second-order behavioral effects that would identify complete tax incidence in this sense. Such effects depend on highly context-specific elasticities.) Second, these counterfactual studies are fundamentally at odds with the accounting framework we work with, because under these types of models, national income will in general be different on a pretax (or no tax) basis compared to a post-tax basis. DINA series can be used as an input for such models, by providing the relevant parameters, but they are not meant to answer such questions directly.

Direct Taxes For direct taxes, the implication of this framework is straightforward: social contributions are paid by the employees, the personal income tax by the respective income earners, etc. The same is true for certain (so-called "other") direct taxes on production, such as property taxes, which are attributed to whoever owns the corresponding property.

Corporate Taxes The corporate tax is paid on business profits. We assume that undistributed profits belong to the shareholders of the corresponding companies. Thus, the corporate tax is paid by those same shareholders.

Indirect Taxes For indirect taxes, especially sales taxes and VATs (i.e., the majority of taxes on production), the implications are more subtle, and to add indirect taxes to a pretax income distribution is more delicate. In essence, the VAT acts as the wedge between factor prices and market prices: therefore, its direct, mechanical effect is on prices.[15] The question then becomes: what is the distributional impact of price changes? Here, we believe it is useful to make a distinction between the distribution of indirect taxes in pretax income and their distribution when moving from pretax to post-tax income.

Let us start with pretax income. Factor price national income (national income excluding indirect taxes) can buy the full production at pretax prices (prices received by producers that do

[15] Indeed, let us first follow the convention of measuring total value added at market prices (i.e., including VAT). If we follow this convention, then GDP is equal to the sum of all the value added in the economy. A firm's value added is measured as the difference between its production and intermediate consumption. But the value of a firm's output is measured at its selling price, which includes VATs: a fall in VATs therefore mechanically leads to a fall in prices. This has the effect of lowering nominal GDP. Conversely, GDP in volume is always calculated on the basis of prices prior to the VATs cut, and therefore remains unchanged. The VATs cut therefore has the effect of lowering nominal GDP without changing real GDP: in other words, it reduces the GDP deflator.

We reach the exact same conclusion if we follow the SNA convention of measuring value added at factor prices. In this case, GDP is equal to the sum of value added and taxes on products. By construction, the VATs is excluded from value added, so its mechanical impact on value added is zero. A reduction in VATs therefore reduces the value of taxes on products without changing the value added, which reduces nominal GDP. What about real GDP? The amount of VATs in volume is calculated by national accounts statisticians by applying the prices and the VATs rate before the decrease to volumes after the decrease. In other words, the VATs cut does not affect GDP in volume terms. The mechanical impact of VATs is, once again, only observed on the deflator, i.e., prices.

not include indirect taxes). Market price national income (national income including indirect taxes) can buy the entire production at post-tax prices (prices paid by consumers, which include indirect taxes). In national accounts, prices are always measured post-tax (i.e., including VATs, sales taxes, etc.) which is why standard national income includes indirect taxes. Factor price income cannot buy full production at post-tax prices precisely because indirect taxes create a wedge between pretax and post-tax prices. Therefore, for pretax national income, factor incomes should be inflated uniformly to line up with the national income aggregate. That way, they reflect the purchasing power of pretax income at the post-tax prices that exist in the economy. Because this is pretax (before any consumption decision is made), it makes the most sense to do a uniform rescaling, so as to preserve the same distribution as factor income. In other words, going from factor price to market price national income is about changing the price index, and not about distributing taxes to individuals.

The issue is different when we consider moving from pretax income to post-tax income. At this stage, we are effectively removing the amount of tax paid by each individual from their income, so it makes sense to ask the amount of sales or VAT paid by each specific person. For that, we favor the view that these taxes are taxes on consumption, and should therefore be assigned to specific consumers. Concretely, in absence of direct data on consumption and savings, indirect taxes would be assigned based on disposable cash income minus saving (where saving rates are set by income groups based on external evidence and are typically growing with income). When more precise data is available, one should consider how different baskets of goods and services that are consumed at different income levels are taxed differently (because of preferential tax rates, e.g., on basic necessities).

Therefore, to summarize, benchmark DINA series distribute basic-price national income (i.e., national income exclusive of indirect taxes) in both pre-tax and post-tax series. Indirect taxes are not included in pre-tax income and are subtracted in post-tax series proportionally to consumption. In a final step, we rescale to market-price national income (i.e. full national income which is inclusive of indirect taxes) proportionally to leave the distributions unchanged. The rescaling to full national income is done for international comparisons as international purchasing power parities are always based on market-prices (inclusive of indirect taxes).

2.2.2.3 Distribution of the Property Income of the Government

For the property income of the government, we recommend the simplest solution, which is to allocate it in proportion to the distribution of factor income, as a level shifter. We could also think of more sophisticated rules, such as an imputation in proportion to taxes paid and benefits received, or different imputation rules for public and private pension surpluses (see Piketty, Saez, and Zucman (2018) for a more detailed discussion in the case of the United States). Note that the same approach applies to the public share of corporate retained earnings (see section 2.2.1.1), which effectively acts as capital income for the government.

2.2.3 Pretax, Post-replacement Income

Pretax, post-replacement national income, which for simplicity we often refer to as "pretax income," is equal to the sum of all pretax income flows accruing to the individual owners of the production factors (labor and capital) before the operation of the tax and transfer system, but after the operation of the social insurance system.

The central difference between pretax factor income and pretax national income is the treatment of pensions, which are counted on a contribution basis by pretax factor income and on a distribution basis by pretax national income. The key reason why we tend to prefer the "pretax national income" series is that it is less sensitive than pretax factor income inequality to the age structure of the population. As discussed in section 2.2.2, pretax factor income inequality is artificially large in economies with a large retired population (even if the pension system is fully contributory, with no redistributive component). In contrast, we aim to define pretax national income so as to satisfy the following neutrality condition: in a hypothetical steady-state 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 of the two), the cross-sectional inequality of pretax 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 overlapping generation 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 issues surrounding the distribution of taxes, retained earnings and government property income are the same for pretax, post-replacement income as for pretax factor income, so for them we refer to section 2.2.2. The first key question for pretax, post-replacement income is what counts as a replacement income. For that we introduce two variants: broad (i.e., including all contributions and benefits) and narrow (i.e., including only pensions) pretax incomes. The second key question concerns the surplus or deficit of the social insurance system.

2.2.3.1 Broad vs. Pension-Based

In the "broad" definition, we deduct all social contributions (D61, see table 2.4). And we add all the social insurance benefits (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 are treated as "replacement income" together with pensions, and which usually represent taxable income in the eyes of the tax authorities).

In the "pension-based" definition, by contrast, we deduct from factor income pension contributions (the sum of D6111, D6121, D6131, D6141, which includes contributions made by employers and households to public and private pension systems, see table 2.4), as well as the investment income payable to pension entitlements (D442). Then we add pension benefits (the sum of D6211 and D6221).

Note that SNA 2008 distinguishes between social *insurance* benefits (the sum of D621 and D622) 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, and include them solely in post-tax income.

If social assistance benefits are excluded from pretax income, 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 full detailed classifications defined by the SNA. Typically, the decomposition between pension and non-pension social insurance benefits and contributions is not available. When it is available, or when other data sources allow the decomposition, we recommend applying both definitions of pretax income, and to compare the level and trends in inequality.

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 that we neutralize the impact of "old age risk" by including pensions. In a number of countries, unemployment insurance benefits are approximately proportional to contributions, like pensions, in which case it makes sense to treat them together. More generally, the "broad" definition aims to include all forms of "social insurance income" (or "replacement income") into pretax 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 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, we would recommend to exclude family benefits from “social security benefits in cash” (D621) and treating them as part of “social assistance benefits in cash” (D623), on the basis that family benefits are non-contributory (that is, they usually bear little relation with contributions to the family benefit funding mechanism).

2.2.3.2 Surplus or Deficit of Social Insurance Systems

Another issue in the pretax, post-replacement income concept 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 2.4). In some countries one might observe a major imbalance between contributions and benefits, partly for legitimate economic reasons (e.g., it could be that the pension system is temporarily accumulating surpluses or deficits, due to demographic changes), and partly for spurious 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 expenditures that are not counted as social insurance benefits). In both cases, we include this surplus (positive or negative) in our definition of pretax national income, so that aggregate pretax national income remains exactly equal to aggregate national income. We recommend treating this deficit or surplus of the pension system like government property income (see section 2.2.2.3). Note also that, to the extent that the corporate sector is responsible for a significant share of the social insurance system (see section 2.1.2.2), and that it experiences a surplus or a deficit, that surplus or deficit must be added to retained earnings in pretax, post-replacement income.

When a deficit or a surplus appears because of our own recategorization of social security benefits (for example, because we have categorized family benefits as social assistance in cash, see section 2.2.3.1) then we must also reclassify the social contributions that pay for them, and try to the extent possible to define social contributions and social benefits in a consistent manner.

In some countries (e.g., Denmark), social contributions are virtually non-existent because the social insurance system is primarily financed using other forms of tax revenue, which generate an artificially large “deficit” of the social insurance system. In such cases, it makes sense to recast part of the direct taxes as social contributions (see Blanchet, Chancel, and Gethin, 2022).

It is worth noting that even if the pension system is at a steady state, and even in the absence of any artificial 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 given by $g = n + h$ (the sum of demographic and productivity growth), and an average return to capital r. With a pay-as-you-go pension system (today's workers pay what today's retirees receive), contributions are by definition equal to pensions, so there is no surplus. However with a funded pension system (today's retirees' receipts are a strict function of their own earlier contributions, regardless of today's workers' contributions) with total steady-state pension wealth equal to $W_t^P = \beta^P Y_t$ (where Y_t is national income, growing at rate g, W_t^P is the 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_t^P. 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 (2018), while it is close to 0% in France (Garbinti, Goupille-Lebret, and Piketty, 2018).

There are all sorts of out-of-steady-state reasons why a pension system may be imbalanced, 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 pretax national income. Note, however, that in the case of steady-state surplus within a funded pension system (as compared to a pay-as-you-go pension system, with no surplus by definition), it is not entirely clear whether one should include that surplus in national income.[16]

[16] 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 a country with a pension fund will accumulate more wealth and will therefore have a national income that exceeds that of a country with a pay-as-you-go system by rW_t^P. However, the pension-fund country needs to save an extra amount equal to gW_t^P 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_t^P (while in the pay-as-you-go country there is no such steady-state pension surplus: contributions are equal to benefits). In other words, national income is higher in the pension-fund country, by rW_t^P, but if we deduct these extra savings the real difference in terms of steady-state resources available for consumption and investment is $(r - g)W_t^P$. As is well known, funded pensions make sense only when the dynamic efficiency condition $r > g$ is satisfied (otherwise we are already in a situation of excessive capital accumulation).

2.2.4 Post-Tax Disposable Income

Post-tax disposable income is defined as pretax factor income (see section 2.2.2), minus indirect taxes (see section 2.2.2.2), taxes on income and wealth (D5, S13), plus social assistance benefits in cash (D623, S14). Social assistance benefits in cash are naturally distributed to their recipients. Retained earnings are still distributed to individuals, net of corporate tax. Note that we do not redistribute all government spending here (see section 2.2.5), so this aggregate does not sum up to national income.

2.2.5 Post-Tax National Income

Post-tax national income is equal to the net national income aggregate. To arrive at the post-tax national income concept, we take post-tax disposable income and add the remaining government spending, i.e., individual and collective consumption expenditure of the government (P3, S13), and the surplus or deficit of the government (see table 2.8), such that it adds up to national income.

2.2.5.1 Allocation of Government Spending

Government spending is the sum of all expenditure made by the general government, including both the central government and local governments. It includes all cash transfers and quasi-cash transfers, such as food stamps, in-kind transfers such as education and healthcare, and collective government expenditure on public goods such as roads, administration, and defense. See Gethin (2024a) for conceptual definitions and details on the construction of government expenditure series in the WID.

We should make clear that it is extremely difficult to allocate certain types of government spending to individuals, and it is not even clear how to attribute public spending such as roads or police to individuals. One important advantage of doing so 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. Another advantage is that it ensures that post-tax income measures are balanced, in the sense that the amount of taxes and transfers distributed to individuals are quantitatively comparable. This is particularly relevant in low- and middle-income countries, where the bulk of redistribution is made in the form of in-kind transfers and public goods. As a result, solely adding cash transfers to individual incomes but removing all taxes would lead to posttax income measures that are dramatically lower than pretax income in these countries.

The DINA methodology considers three alternative methods for allocating in-kind transfers to individuals, each with advantages and disadvantages. A first possibility is to allocate these transfers proportionally to disposable income. The main advantage of this methodology is that it is simple, transparent, and neutral. By doing so, we simply raise all income levels and do not change the distribution. The main disadvantage is that it treats all types of public services in the same way, while some transfers are arguably more progressive than others (e.g., shelters for homeless people versus subsidies to multinational corporations). It also implies that more unequal countries have a proportionally more unequal distribution of public goods, which may not always be true empirically.

A second possibility is to use a lump-sum method: attribute the average monetary value of all in-kind transfers and collective expenditure to each adult individual. This approach also has the advantage of being simple and transparent, but it also ignores the fact that some transfers are more progressive than others. Furthermore, while a lump sum allocation might be justified for certain expenditures that arguably benefit everyone equally, such as national defense or justice, many public services are likely to benefit some groups more than others. For example, in education, children from wealthier families are more likely to access tertiary education, and spending on primary education can vary greatly depending on municipal budgets if financed through local taxes — potentially reproducing or deepening geographical inequalities. At the same time, some education spending is often designed to offset these disparities, such as preschool programs targeting children from low-income families. Additionally, poorer populations often experience higher rates of illness and rely more heavily on public health services. Other public services can also pull in different directions. Subsidized social housing or primary health centers in rural areas can disproportionately benefit lower-income groups, while spending on infrastructure such as highways or cultural institutions may favor wealthier urban centers.

A third possibility is thus to explicitly attempt to estimate the distributional incidence of specific public services. The main public services that have been studied in the literature are healthcare and education. Healthcare is the public service on which we have data of the highest quality. The general conclusion is that it is highly progressive, often close to a lump sum allocation or even more egalitarian. Education has also been found to be highly progressive, mainly because poorer households have more children and disproportionately rely on public (rather than private) schools.[17] A major caveat here that we lack sufficiently granular data on spatial inequalities in

[17] On healthcare, see for instance Piketty, Saez, and Zucman (2018) on the United States, Bruil et al. (2025) on the Netherlands, Germain, André, and Blanchet (2021) on France, and Gethin (2024c) on South Africa. On education, see for instance Gethin (2024c) on South Africa and Riedel and Stichnoth (2022) on the United States. Among other efforts, the Commitment to Equity database contains estimates of the distributional incidence of education

public education spending, implying that the progressivity of education is likely overestimated in these studies. Evidence on other public services is extremely limited, although there is some evidence that their distribution may fall in-between a lump sum and proportional allocation (Currier, Glaeser, and Kreindler, 2024; Gethin, 2024c).

Given these large uncertainties, the WID publishes post-tax income series that currently make the conservative assumption of proportionality for all government consumption and other in-kind transfers than healthcare. We make an exception for healthcare given the large empirical evidence on its progressive nature: we distribute it on the basis of surveys reporting information on intensity of use of the public healthcare system, often implying that its allocation is close to a lump sum (M. Fisher-Post and Gethin, 2023; Gethin, 2024b). Hence, WID benchmark series assume a close to lump-sum imputation of healthcare expenditure and a proportional imputation of all other items. In the future, we hope to publish alternative series using different assumptions, such a lump sum allocation for all non-health expenditure or other alternatives based on available data.

2.2.5.2 Allocation of Government Surplus or Deficit

We treat the surplus or deficit of the government like government property income (see section 2.2.2.3). Note that our notion of government deficit or surplus differs slightly from the ones that are usually discussed (see section 2.1.2.3).

2.2.6 The Female Labor Income Share

Since 2021, based on the work done by Neef and Robilliard (2021) for the 2022 World Inequality Report, the WIL database includes a series of country-level estimates of the female labor income share (FLIS). Alongside vertical inequality captured by the income distribution series presented above, the objective of FLIS is to capture an important aspect of group-level inequality.

Gender inequality in employment and labor income is often assessed through two distinct lenses: access to employment and earnings levels. However, both dimensions matter from an income inequality perspective since total labor income earned by women at the national level depends on both the employment rate and average earnings. The Female Labor Income Share accounts for both aspects by capturing the share of total labor income earned by women within a country. It is calculated as the ratio of women's labor income to the aggregate national labor income:

and healthcare in many developing countries (see Lustig, 2018). See also Gethin (2024b) and M. Fisher-Post and Gethin (2023) for attempts at estimating the distributional incidence of education and healthcare at the global level.

$$\text{Female Labor Income Share} = \frac{\text{Labor income received by women}}{\text{Total labor income}}$$

Labor income includes wage and salary income as well as the labor share of self-employed income:

$$\text{Labor income} = \text{Wage income} + \text{The labor share of self employment income}$$

To compute the female labor income share, labor income is aggregated by gender within each country. More detail on the methodology is provided in chapter 5. Unlike the gender wage gap—which measures differences in pay between women and men in comparable positions—the female labor income share captures both wage differentials and disparities in employment. As a result, it reflects women's overall share of national labor income and is generally lower than the gender pay ratio, since it combines the effects of lower average earnings with lower labor force participation.

Key Points

- DINA income series distribute the entirety of net national income using concepts that are consistent with the SNA. Therefore, they include certain types of income (like the undistributed profits of corporations or indirect taxes) that are traditionally overlooked by other sources.

- We define four broad types of series. Pretax factor corresponds to the distribution of income before any redistribution, be it through social insurance systems of social assistance. Pretax post-replacement income includes redistribution that occurs through social insurance schemes (pension and sometimes unemployment benefits), but not social assistance. Post-tax disposable income includes all cash redistribution through the tax and transfer system, but does not include in-kind benefits and therefore does not add up to national income. Finally, post-tax national income redistributes all in-kind transfers (i.e., government consumption expenditures) to individuals.

- Because retirees generally have very little pretax factor income, that concept is very sensitive to the age structure of the population. This is why we prefer to use pretax post-replacement income as our benchmark pretax income concept.

- The undistributed profits of corporations are distributed to the shareholders of these corporations. Direct taxes and benefits are distributed to the people that pay the tax or receive the benefit. The corporate tax is paid by the shareholders. Sales taxes and VATs are distributed proportionally to factor income for pretax concepts, and are removed proportionally to consumption in post-tax concepts.

- For in-kind transfers, we consider three variants: allocation proportional to post-tax disposable income, lump-sum allocation, and proportional allocation except for health spending which we distribute lump sum. We use that last concept as our benchmark for post-tax national income.

Chapter 3

Wealth Concepts

This chapter defines the various concepts of wealth, assets and rates of return that we use in the WID. Like the income concepts in chapter 2, our wealth concepts are defined using the 2008 SNA. We only deviate from these concepts in the treatment of unfunded employers' pensions.

We again illustrate the concepts using data from the "Sequence of Accounts" provided alongside the SNA 2008 guidelines (United Nations, 2009). Because this "Sequence of Accounts" does not provide a sufficiently detailed breakdown of certain nonfinancial assets, we complete it using data from France's "Table of Integrated Economic Accounts" for the year 2018, produced by the French statistical institute. The online appendix provides details of the computations. 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 in chapter 6). Finally, we present the definitions of national and public wealth that we use in the WID.

As a rule, all financial assets and liabilities of resident sectors (i.e., households, nonprofit institutions serving households, corporations, general government) are unconsolidated.[1] For the rest of the world, series are consolidated. For the public sector, we separately report consolidated liabilities, in addition to the unconsolidated ones.

[1]The SNA guidelines indicate that "the accounting entries in the System are not consolidated. Therefore, the financial balance sheet of a resident sector or subsector is to be presented on a non-consolidated basis" (United Nations, 2009).

3.1 Personal and Nonprofit Wealth

3.1.1 Stock of Wealth

3.1.1.1 Decomposition of the Stock of 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. The details of the computations are given in table 3.1, where we also provide a number of decompositions into different classes of assets.

	Gross personal wealth	4689
AN, S14	Non-financial assets owned by households	1429
	Housing assets of households	*1174*
AN111, S14	*Dwellings owned by households*	*681*
AN21111, S14	*Land underlying dwellings owned by households*	*493*
	Business and other non-financial assets of households	*255*
AN2112, S14	*Agricultural land of households*	*23*
	Other domestic capital of households	*232*
AF, S14	Financial assets owned by households	3260
AF2+AF3+AF4+AF7+AF8, S14	*Currency, deposits, bonds, loans, derivatives and other financial assets of households*	*1120*
AF5, S14	*Equity and investment fund shares of households*	*1749*
AF6, S14	*Life insurance and pension funds of households*	*391*
AF, S14	Minus: Liabilities of households	189
	Equals: Net personal wealth	**4500**

Adapted from the SNA 2008 "Sequence of accounts" (United Nations, 2009) and the French Table of Integrated Economic Accounts (INSEE, 2018). See online appendix for the construction of the table.

Table 3.1: Net Personal Wealth

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 (as in observed real estate transactions) than the two components separately. Business assets (and other non-financial assets) are the difference between total non-financial assets and housing assets. They include items such as fixed assets (non-residential real estate, machinery and equipment), intangible capital or natural resources.[2]

[2]For certain asset types, the difference between "business assets" (a non-financial asset) or "equity" (a financial asset) is rather fuzzy and depends on whether assets are incorporated or not according to specific legal environments. We refer the interested reader to Piketty and Zucman (2014, Appendix, p. 26). This fuzziness is essentially unimportant when it comes to the aggregate value of household wealth, but it does "*affect the comparison of the structure of production across countries, the computation of labor and capital shares, and the analysis of the*

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 the SNA 2008, land appears as a single asset (classification code AN211). In the detailed ESA 2010 classification codes, land (AN211) is broken down into "land underlying buildings and structures" (AN2111), "land under cultivation" (AN2112), "recreational land and associated water surfaces" (AN2113), and "other land and associated water surfaces" (AN2114). Many national statistical agencies also break down "land underlying buildings and structures" (AN2111) into "land underlying dwellings" (AN21111) and "other land underlying buildings and structures" (AN21119). When this latter decomposition is not available, we recommend splitting the land value in proportion to value of dwellings and other buildings and structures.

Whenever possible, we also recommend breaking down business assets (and other non-financial assets) into agricultural land (AN2112) and other domestic capital (i.e., all non-financial assets except housing and agricultural land). Although agricultural land is now a negligible part of assets in the balance sheet of developed countries, it played a very large historical role, and still plays an important role in developing countries with a large agricultural sector. 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 deserves attention.[3] We include natural capital (AN212) – such as mineral and energy reserves and which can be very substantial in certain countries, in particular in developing ones – under our residual category "Other domestic capital" since few countries actually report these data. The aim of the WID is to progressively account for these assets.

Finally, we split financial assets into three categories: currency, deposits, bonds, loans, derivatives and other financial assets (the sum of AF1[4], AF2, AF3, AF4, AF7 and AF8), equity and investment fund shares (AF5), and life insurance and pension funds (AF6). 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

structure of household wealth" (*ibid*).

[3]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 the SNA 2008 and the ESA 2010, whenever it is impossible to separate land and building value, all value is allocated to the biggest part. Also, note that "other buildings and structures" (AN112) are broken down into "buildings other than dwellings" (AN1121), "other structures" (AN1122), and "land improvement" (AN1123). However AN1123 — 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 and Zucman (2014) and Piketty (2014, chap. 6).

[4]AF1 "Monetary gold and special drawing rights" is usually very small and only held by Central Banks, which is why it is not displayed here in the explanatory tables.

capitalization method.

The WID deviates from the treatment given by the SNA 2008 to unfunded employers' pensions. Namely, the SNA 2008 treats unfunded employers' pensions as wealth, and the associated investment income flow as part of investment income payable on pension entitlements (as such, this income would be part of our pretax pension income). In our view the SNA treatment is not satisfactory. Unfunded private pensions are not wealth. They are promises of future transfers that are not backed by actual wealth. In practice, in the United States, 99% (in 2018) of unfunded pension entitlements are for state and local government workers. These promises are thus conceptually similar to promises of future Social Security payments, which are not treated as wealth. Similarly, it is unsatisfactory to treat the investment income "earned" by unfunded pension plans as pretax income, since there is no such income. This income should be removed from "investment income payable on pension entitlements" and added to corporate saving. It is conceptually inconsistent to treat unfunded pensions as wealth if one does not treat promises of future Social Security payments as wealth. And it is inconsistent to treat promises of future Social Security payments as wealth if one does not treat other promises of future government transfers (such as promises of future health spending, spending for one's children, etc., net of future taxes) as wealth — a computation that is both impractical (how to value these "assets"?) and conceptually flawed (these "assets" are not wealth, they cannot be sold on a market, they have no market value, nobody can claim them as their property). In the United States, Saez and Zucman (2016) remove unfunded pensions from wealth (and Piketty, Saez, and Zucman (2018) remove the corresponding income flows from pension income).

Note that in some countries, available balance sheets include the assets and liabilities of the nonprofit sector together with those of the household sector. In such cases, we cannot compute personal wealth, and we can only compute private wealth (as defined by the sum of personal wealth and nonprofit wealth). Given that nonprofit wealth can represent a non-negligible fraction of private wealth, we recommend to estimate at least some approximate breakdown of private wealth into personal and nonprofit wealth, for instance by using the decomposition of capital income flows (which are more often available separately for the household and nonprofit sectors, though not always). When balance sheets are available separately for both the household sector and the nonprofit sector, then we can easily account for nonprofit wealth using the same decomposition as for personal wealth (see table 3.2), and we can compute private wealth as the sum of the two.

	Gross non-profit wealth	331
AN, S15	Non-financial assets owned by nonprofits	159
	Housing assets of nonprofits	*0*
AN111, S15	*Dwellings owned by nonprofits*	*0*
A21111, S15	*Land underlying dwellings owned by nonprofits*	*0*
	Business and other non-financial assets of nonprofits	*159*
AN2112, S15	*Agricultural land of nonprofits*	*0*
	Other domestic capital of nonprofits	*159*
AF, S15	Financial assets owned by nonprofits	172
AF2+AF3+AF4+AF7+AF8, S15	*Currency, deposits, bonds, loans, derivatives and other financial assets of nonprofits*	*146*
AF5, S15	*Equity and investment fund shares of nonprofits*	*22*
AF6, S15	*Life insurance and pension funds of nonprofits*	*4*
AF, S15	Minus: Liabilities of nonprofits	121
	Equal: Net non-profit wealth	**210**

Adapted from the SNA 2008 "Sequence of accounts" (United Nations, 2009) and the French Table of Integrated Economic Accounts (INSEE, 2018). See online appendix for the construction of the table.

Table 3.2: Net Nonprofit Wealth

3.1.1.2 Wealth vs. Capital

Some words on the conceptual difference between "wealth" and "capital" are warranted at this stage. The difference can be directly traced from the sequence of non-financial accounts: profit (operating surplus and mixed income) is the flow of income accruing to the owners of the produced capital stock necessary for production, while economic rent (property income) is the flow of income accruing to the owners of financial assets (equity, bonds, derivatives, etc.) or tangible non-produced assets (land or subsoil assets).

Wealth is thus a more encompassing term as it not only includes assets produced by humans (the "capital stock" of buildings, equipment, infrastructure) but also includes assets that are not productively employed to generate new income (financial assets, natural resources). The broad definition of wealth follows the pure accounting and monetary definition that is used in balance sheets by businesses. It thus corresponds to the total amount of funds invested in enterprise, embodied in the monetary value of tangible and intangible assets, and usable as collateral to obtain credit to expand production (Hodgson, 2014).

Such a monetary definition of wealth is subject to the vicissitudes of financial market valuations, which can make the overall value of the stock of wealth diverge substantially from the value of the capital stock as measured by its replacement cost (i.e "book-value"). The latter is the concept that economists commonly use in research on economic growth, given its fundamental link to GDP, which is concerned with the expansion of produced goods and services. Our interest in market-value wealth is motivated by the distributional concerns of DINA. The market value of wealth can better approximate the micro distribution of wealth, as well as the resulting income flows that are actually available for consumption and saving. Moreover, as highlighted above, growing market values expand the credit opportunities for businesses to expand their productive operations. Thus, we deem it important to report market-value wealth as well as the capital stock at book value.

3.1.2 Rates of Return

We provide in table 3.3 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 for the personal sector. These average flow rates of return exclude capital gains and are computed by linking the classification of assets with 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 chapter 6).

income code	wealth code	type of asset	income	wealth	rate of return
B2n (S14)	AN111 +A21111 (S14)	housing	77	1174	6.6%
capital share (40%) of B3n + net D45 (S14)		business	11	255	4.3%
D4 excluding D45 (S14) + private share of B5n (S11+S12)	AF (S14)	financial assets	238	3260	7.3%
D4 excluding D45	AF (S14)	financial liabilities	16	189	8.3%

Adapted from the SNA 2008 “Sequence of accounts” (United Nations, 2009) and the French Table of Integrated Economic Accounts (INSEE, 2018). See online appendix for the construction of the table.

Table 3.3: Rates of Return on Personal Wealth

In table 3.3 we present rates of return for four assets and liabilities: housing assets, business (and other non-financial) assets, financial assets, and liabilities. These are pretax rates of returns, and we attribute undistributed profits (including corporate income tax payments) to financial assets. We also attribute all production taxes to factor labor and capital income flows in proportion of each income flow. This is be acceptable as a first approximation, but it could be improved: for example, property taxes could be attributed to housing (see the discussion in chapter 5). Note that the WID provides separate series for production taxes, so that users can make their own imputations if they want to.

income code/formula	wealth code/formula	type of asset	income	wealth	rate of return
		undistributed profits to equity			
D41 (S14)	AF2+AF3+AF4 +AF7+AF8 (S14)	currency, deposits and debt assets	49	1120	4.4%
D42+D43 (S14) and B5n (S11+S12)	AF5 (S14)	equity and investment fund shares	135	1749	7.7%
D44 (S14)	AF6 (S14)	life insurance and pension funds	30	391	7.7%
		undisributed profits to equity and life insurance			
D41 (S14)	AF2+AF3+AF4 +AF7+AF8 (S14)	currency, deposits and debt assets	49	1120	4.4%
D42+D43 (S14) and part of B5n (S11+S12)	AF5 (S14)	equity and investment fund shares	72	1749	4.1%
D44 (S14) and part of B5n (S11+S12)	AF6 (S14)	life insurance and pension funds	93	391	23.9%
		undisributed profits to all financial assets			
D41 (S14) and part of B5n (S11+S12)	AF2+AF3+AF4 +AF7+AF8 (S14)	currency, deposits and debt assets	103	1120	9.2%
D42+D43 (S14) and part of B5n (S11+S12)	AF5 (S14)	equity and investment fund shares	48	1749	2.8%
D44 (S14) and part of B5n (S11+S12)	AF6 (S14)	life insurance and pension funds	63	391	16.1%

Adapted from the SNA 2008 "Sequence of accounts" (United Nations, 2009) and the French Table of Integrated Economic Accounts (INSEE, 2018). See online appendix for the construction of the table.

Table 3.4: Rates of Return on Financial Wealth

Table 3.4 breaks 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, investment fund shares, life insurance and pension funds, or the sum of all financial assets. In countries such as the United States, where there exists information on the composition of the wealth of pension funds and life insurance companies (especially 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 and life insurance companies' wealth which is invested in equities.

Finally, note that the transition from the SNA 1993 to the 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 stayed virtually unchanged in the new system, so that we can define our three main categories of financial assets in the same manner: deposits, currency, bonds and loans (sum of AF1, AF2, AF3, AF4, AF7, AF8), equity and mutual funds (AF5), life insurance and pension funds (AF6). 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 in D42 (together with dividends and other property income flows going to AF5 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-Lebret, and Piketty (2020) in the case of France). In some cases, we may also prefer to isolate "deposits and currency" (AF2) within the broader asset category "deposits, currency, bonds and loans" (see also Garbinti, Goupille-Lebret, and Piketty, 2020). In all cases, we recommend to perform multiple sensitivity tests when applying the income capitalization method (see discussion in chapter 6 and in the country-specific papers).

3.2 Private and Public Wealth

3.2.1 Private Wealth

Private wealth is the net (assets minus liabilities) wealth of households and NPISH. It can therefore be broken down similarly to personal and nonprofit wealth, as presented in section 3.1.1. This decomposition is shown in table 3.5, which adds up tables 3.1 and 3.2.

3.2.2 Public Wealth

Public wealth is similarly defined as the net (assets minus liabilities) wealth of the general government sector. It can be broken down similarly to private wealth. Table 3.6 presents this decomposition.

Note that we provide *unconsolidated* public debt by default, which means that our numbers can differ from those discussed in other contexts. In previous versions of the World Inequality Database, we also provided *consolidated* series for public debt (*gwdec*) in addition to standard unconsolidated series (*wgdeb*). However, given the limitations of available data sources, it seems very difficult at this stage to estimate consolidated vs unconsolidated public debt at the global level. We therefore decided to concentrate entirely on unconsolidated series for the time being.

	Gross private wealth	5020
AN, S14+15	Non-financial assets owned by the private sector	1588
	Housing assets of the private sector	*1174*
AN111, S14+15	*Dwellings owned by the private sector*	*681*
A21111, S14+15	*Land underlying dwellings owned by the private sector*	*493*
	Business and other non-financial assets of the private sector	*414*
AN2112, S14+15	*Agricultural land of the private sector*	*23*
	Other domestic capital of the private sector	*391*
AF, S14+15	Financial assets owned by the private sector	3432
AF2+AF3+AF4+AF7+AF8, S14+15	*Currency, deposits, bonds, loans, derivatives and other financial assets of the private sector*	*1266*
AF5, S14+15	*Equity and investment fund shares of the private sector*	*1771*
AF6, S14+15	*Life insurance and pension funds of the private sector*	*395*
AF, S14+15	Minus: Liabilities of the private sector	310
	Equal: Net private wealth	**4710**

Adapted from the SNA 2008 "Sequence of accounts" (United Nations, 2009) and the French Table of Integrated Economic Accounts (INSEE, 2018). See online appendix for the construction of the table.

Table 3.5: Net Private Wealth

More research is needed in order to produce reliable estimates of unconsolidated vs consolidated balance sheets, whether this is at the level of the government sector, the corporate sector, the household sector or each country or region as a whole.

3.3 National, Corporate and Foreign Wealth

3.3.1 Corporate Wealth

There are two ways of computing the value of corporations. We can use their market value (i.e., their equity liability) or we can use their book value (i.e., the difference between their assets and their non-equity liabilities). The book value of corporations can be broken down similarly to public and private wealth. This is presented in table 3.7. This table also includes the market value of corporations as a memorandum item.

The difference between the market value and the book value of corporations is what we call the "residual corporate wealth." We can also define Tobin's Q as the ratio between the market value and the book value (i.e., $2752/1928 \approx 1.4$).

Residual corporate wealth 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

	Gross government wealth	1099
AN, S13	Non-financial assets owned by the government	783
	Housing assets of the government	*44*
AN111, S13	*Dwellings owned by the government*	*25*
A21111, S13	*Land underlying dwellings owned by the government*	*19*
	Business and other non-financial assets of the government	*739*
AN2112, S13	*Agricultural land of the government*	*0*
	Other domestic capital of the government	*739*
AF, S13	Financial assets owned by the government	316
AF2+AF3+AF4+AF7+AF8, S13	*Currency, deposits, bonds, loans, derivatives and other financial assets of the government*	*284*
AF5, S13	*Equity and investment fund shares of the government*	*12*
AF6, S13	*Life insurance and pension funds of the government*	*20*
AF, S13	Minus: Liabilities of the government	687
	Equal: Net government wealth	**412**

Adapted from the SNA 2008 "Sequence of accounts" (United Nations, 2009) and the French Table of Integrated Economic Accounts (INSEE, 2018). See online appendix for the construction of the table.

Table 3.6: Net Public Wealth

measured by their market equity value — is already included in the financial assets and therefore the net wealth of these other sectors.

If Q 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.

When Q is less than one, which is often the case in practice (e.g., in Germany, Japan or France, as well as in the United Kingdom and the United States until the 1990s and the 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, when Q is higher than one, which happens for certain periods and countries (e.g., in the United Kingdom and the United States in the 1990s and 2000s, at least prior to the 2008 financial 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.

	Gross book value of corporations	5941
AN, S11+12	Non-financial assets owned by corporations	2228
	Housing assets of corporations	*700*
AN111, S11+12	*Dwellings owned by corporations*	*380*
A21111, S11+12	*Land underlying dwellings owned by corporations*	*321*
	Business and other non-financial assets of corporations	*1528*
AN2112, S11+12	*Agricultural land of corporations*	*211*
	Other domestic capital of corporations	*1317*
AF, S11+12	Financial assets owned by corporations	3713
AF2+AF3+AF4+AF7+AF8, S11+12	*Currency, deposits, bonds, loans, derivatives and other financial assets of corporations*	*2827*
AF5, S11+12	*Equity and investment fund shares of corporations*	*831*
AF6, S11+12	*Life insurance and pension funds of corporations*	*55*
AF, S14	Minus: (Non-equity) liabilities of corporations	4013
	Equal: Net book value of corporations	**1928**
	Memo: Equity liability of corporations (market value)	2752

Adapted from the SNA 2008 "Sequence of accounts" (United Nations, 2009) and the French Table of Integrated Economic Accounts (INSEE, 2018). See online appendix for the construction of the table.

Table 3.7: Net Book Value of Corporations

3.3.2 National Wealth (Book vs. Market Value)

The distinction between the book value and the market value of corporations means that we can give two different, complementary definitions for national wealth.

First, we define market-value national wealth as the sum of private wealth and public wealth (i.e., we ignore residual corporate wealth, which cannot be directly attributed either to private individuals or to the government).[5] In that definition, the market value of corporations is indirectly accounted for through the ownership of corporations by both the public and the private sector. Second, we define book-value national wealth as the sum of market-value national wealth and the residual wealth of corporations.

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

[5]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 nonprofit 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).

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 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 (i.e., the net foreign wealth). This definition can be meaningful if deviations of Tobin's Q from unity do not reflect measurement errors only, but also reflect real changes in the balance of power between the various stakeholders of corporations. A Tobin's Q lower than one (positive residual corporate wealth) might reflect a situation where stakeholders other than shareholders partly control companies' income flows (as in Germany). Conversely, Tobin's Q higher than one (negative residual corporate wealth) might reflect a situation where shareholders are able to extract high rent from companies, maybe because the legal system is very favorable to them (as in the United States and United Kingdom). In these cases, "book-value national wealth" can be interpreted as capturing the value of corporations not from the viewpoint of the firms' owners (as reflected in equity and bond prices, and captured by the market value), 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.

Moreover, both measures of national wealth can be rewritten as the sum of domestic non-financial assets (i.e., agricultural land, housing assets, and other domestic capital) and net foreign wealth (see the definition in the following subsection). When other domestic capital is measured at market value, this definition corresponds to market-value national wealth and when other domestic capital is measured at book value, it corresponds to book-value national wealth (see Piketty and Zucman, 2014, section III.A, p. 1269). Such a definition is possible because net foreign wealth equals the sum of the net domestic financial assets of domestic sectors. Hence, by measuring non-financial assets and adding net foreign assets, one is back to the definition of national wealth as the sum of net wealth of domestic sectors. 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.

Finally, note that although we are primarily interested in estimating the distribution of personal wealth among private individuals (see chapter 6), it could also be interesting in some cases to construct estimates of the distribution of national wealth among individuals. For example, in a country with a large public wealth (such as Norway), it may make sense to attribute public wealth to private individuals. Otherwise, the residents of this country (e.g., 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 the distribution of tax liabilities. Another would be in proportion to the distribution of entitlement spending (e.g., rights to pensions financed out of public sovereign funds). Those different methods of attributing national wealth to individuals might differ substantially in their distributional implications, but this is not necessarily a reason to omit aggregate national wealth from distributional considerations. In countries where sovereign wealth is controlled by a smaller group of the population (e.g., in Gulf countries), one might choose to apply specific imputation methods for that segment of the population. These are important and complex issues that we leave for future research.

3.3.3 The Global consistency of the net foreign wealth

Net Foreign Wealth, or the Net Foreign Asset (NFA) Position, refers to the difference between the value of a country's external financial assets and its external financial liabilities at a given point in time. Equivalently, it can be calculated by summing all financial assets held by domestic sectors (households & NPISH, corporations and the government) and subtracting their total financial liabilities. In a closed economy, these positions net to zero, but in an open economy, the net foreign asset position reflects the net financial claims that residents of a country hold over the rest of the world.

By construction, every asset is someone else's liability. In theory, this implies that the sum of all foreign asset positions should add up to zero, since the world as a whole cannot be in debt. However, in practice, this is not the case and official statistics sum up to a negative world position (Zucman, 2013), which is impossible. To correct for this and to account for hidden wealth offshore, we perform the same proportional method used to adjust foreign income flows (see Section2.1.1.3) to adjust for inconsistencies in stocks. For a detailed discussion, see Nievas and Piketty, 2024.

Key Points

- DINA distributional wealth series distribute the entirety of household wealth to individuals. This includes financial and non-financial assets, minus liabilities.
- For aggregate balance sheets, we decompose the wealth of four sectors: households, NPISH, corporations and the government.
- The value of corporations can be estimated at "market value" (our benchmark) or at "book values" (i.e., the value of their assets, minus their non-equity liability). The difference the market value and the book value of corporations is what we call "residual corporate wealth." It can be positive or negative.
- There are two alternative definitions of national wealth. "Market-value national wealth" measures national wealth using the market value of corporations, while "book-value national wealth" uses their book value.

Part II

Methods

Chapter 4

Macroeconomic Accounts

The WID contains extensive data on income and wealth aggregates across the world compiled from a variety official and academic sources. These aggregates follow the definitions explained in chapters 2 and 3. This chapter explains how these aggregates are constructed. We only explain the general principles. For series-specific details, please refer to the methodological notes associated to each variable, and to country-specific papers. The complete code and data that integrates the various sources together and performs necessary imputations and adjustments described below is available online at `https://github.com/widworld/wid-world`.

4.1 Income, Prices and Populations

The macroeconomic income side of the WID contains the decomposition of national income into the main components that are necessary for the construction of income DINA series.

We provide, at the very least, series of population, net national income and a price index for all the countries in the world since 1950, using some imputations when necessary. These imputations are simple but rough estimates based on limited data. They should be considered provisional, and will be subject to revision whenever better data becomes available.

For the subcomponents of national income, the coverage varies a lot. Since the mid-1990s, official data portals generally provide a reasonable level of coverage for developed countries. We extend this data with work done by researchers for several areas, using country-specific sources and archives, which are usually more complete. This work will have to be extended further before we can achieve a truly satisfying level of coverage.

The construction of the aggregate income series in the database is done in several steps. First, we estimate long-run series of GDP, the price index, and the population. Then we estimate each component of national income as a fraction of GDP, ensuring at every step their plausibility, and their consistency in terms of accounting identities. We perform imputations for the consumption of fixed capital and its sectorial decomposition. Finally, we estimate the additional components of net foreign income (see chapter 2.1.1.1), and distribute them to the appropriate sectors of the economy.

As a result, we are able to provide one of the most complete sources on the decomposition of national income across the world, done in a consistent and harmonized way. This source can be used to study various macro trends, and to help compute DINA income series.

4.1.1 Estimations of Main Aggregates

4.1.1.1 Estimation of Population

We usually define the population of a country as its *de facto* population on the 1st of July of the year indicated. To that end, we use in priority population data provided by the WIL fellows, which usually stems from national demographic institutes. Otherwise, the population series come from the WPP, providing total population, as well as population by age group and by sex, for all countries, since 1950.

But we have to make a few adjustments. For example, the WPP uses the *de jure* definition of countries' boundaries, unlike most national accounts which use *de facto* boundaries (e.g., Cyprus' population in the WPP includes Northern Cyprus, while Cyprus's national accounts exclude it). The same problem arises when countries have experienced border changes: the WPP always uses the contemporary borders, while national account series mostly use the border of the time.

In such cases, we rely on population series published directly by the UN in its AMA database. These series, however, only concern the overall population (including children), while we prefer to focus on the adult population (20 and older). Therefore, in these cases, we rescale the population by age and sex from the WPP using the population total from the AMA, effectively assuming that the population structure by age and sex is similar within the *de jure* and the *de facto* boundaries of countries, or in the former and contemporary borders.

A full description of the construction of the `WID.world` population series, including historical reconstructions back to 1800, projections up to 2100, and consistent age- and gender-specific breakdowns, is provided in Gomez-Carrera et al. (2025)

4.1.1.2 Estimation of GDP

GDP is the only aggregate that we estimate directly when constructing the database. Every other income component is estimated as a proportion of it, and then anchored to the GDP at the end of the procedure. The reason for this approach is that, while we prefer to focus on national income, GDP is the item of national accounts that is the most widely available: that is, while in some cases we know the GDP of a country but not any other component, we virtually never observe a situation in which we know a subcomponent of GDP but not GDP itself. When we combine various sources, we import the components of national income as a fraction of GDP from that same source, as a simple way to get rid of various issues such as inconsistent currencies between sources. If we were to use any other variable than GDP as our denominator, we would lose observations in the process.

Here, again, our priority source is the data sent by WIL fellows, directly collected from the national statistical institutes of countries. If this is not available, we use the series from the UN, the World Bank, the IMF, or Maddison (2007). The UN data is divided in two parts. The MADT part contains highly detailed data on GDP and its subcomponents, going back to 1946 at the earliest. It distinguishes series based on the various versions of the SNA and other secondary methodological aspects. Although rich in information, series from this data source have many breaks. The AMA provides fewer series over a shorter time span (since 1970) but covers the entire period without breaks. The World Bank provides GDP series, usually back to 1990, and sometimes 1960. The IMF data come from its biannual WEO publication. This database only starts in 1980 but provides preliminary estimates and forecasts of GDP, which are very useful to cover the most recent years. Finally, Maddison (2007) provides data of GDP worldwide since the year 0, although we only use its post-1950 estimates. The Maddison (2007) database is used for some of the oldest GDP estimates.

The GDP series are constructed in two steps. First, we pick the GDP level in a given year and from a given source. For countries which have GDP data send by a WIL fellow, we use that GDP level for the most recent year available. Otherwise, we use the most recent data from one of the other sources. In case of conflict, we give priority to the World Bank, then the UN and other sources. We do not use the IMF preliminary estimates or forecasts for fixing the GDP level.

Second, we construct a continuous series of GDP growth rates. As before, we use in priority the data sent by WIL fellows, then the World Bank, then the UN and other sources. If none of these sources has any data, which can be the case for the most recent years, we use the growth rates of the IMF WEO, or as a last resort we carry forward the growth from the last available year. All those sources typically provide data since 1970 (UN), 1960 (World Bank) or 1980 (IMF). For earlier years, we use the real GDP growth rates from Maddison (2007). In some case, we rely on more specific sources. For example, in China, the official GDP growth figures have been subject to criticism. Therefore, we use corrected GDP (and deflator) estimates from Maddison and Wu (2007). Finally, we combine the GDP growth rates with the GDP level to get a unique GDP series covering the entire time period. The series construction for each country (which gives the source used for the level and the growth rate each year) is reported in the methodological notes associated to the series.

4.1.1.3 Estimation of the Price Index

Our estimation of the price index follows the same general principles as our estimation of GDP: that is, we cumulate inflation rates from a variety of sources. The difference is that since price indexes are relative quantities, we do not have to estimate absolute levels. As explained in section 1.2.1, we use GDP deflators (from the same sources as GDP), but fall back on the CPI when these are not available. Again, the series construction for each country (which gives the source used for each year) is reported in the methodological notes associated to the series.

4.1.2 Estimation of National Income and its Components

4.1.2.1 Estimation of Components

On top of the data from WIL fellows, we import national income components as a fraction of GDP from several sources: the UN MADT, the OECD, and the IMF BOPS (for foreign incomes).[1] When doing so, we make several checks and adjustments to the data. This is especially true for the UN MADT database, which a very rich but sometimes less curated source than the OECD or the IMF.

First, we make corrections for categorization issues. For example, many countries do not separate labor and capital incomes from production taxes in their foreign income flows, whereas the UN MADT database structure forces them to, so these countries include production taxes within labor and capital incomes. The UN MADT also has a separate entry for mixed income

[1] Because the IMF BOPS does not provide GDP series directly, we use the GDP series (in current USD) from the UN AMA.

and operating surplus of the household sector, while some countries only report both combined, so they only fill one of these variables. It also asks for both gross operating surplus and gross mixed income, but some countries choose to report, say, net mixed income in the gross mixed income variable, and the sum of net operating surplus and consumption of fixed capital (for both mixed income and operating surplus) in the gross operating surplus variable, or *vice versa*. Inconsistent recording of gross and net values is currently present throughout the database. These issues are usually reported in "footnotes," which we parse to systematically correct them. The OECD database has fewer problems, but one still must be careful: for example, New Zealand inverts uses and resources when recording foreign property income in the primary distribution of income account of the rest of the world. When some numbers lack plausibility and there is no obvious correction, we drop them.

Second, we combine these sources, giving priority to data from WIL fellows, then the IMF, the OECD, and finally the UN MADT. The UN MADT provides series corresponding to different iterations of the SNA methodology, so we give priority to the most recent and up-to-date series.

When combining the series, we want to avoid the apparition of series breaks due to pure methodological differences, so we follow a simple splicing procedure. We calculate the average difference (as a fraction of GDP) between the two series for the same country over the period during which they overlap. Then we correct the lowest-priority series by adding that difference to the entire series. When the series do not overlap, we assume the difference to be zero.

Finally, we have to ensure consistency between all series in terms of accounting identities. Inconsistencies can appear for several reasons. First, they sometimes appear directly in the raw sources. This is what the national accounts call "statistical discrepancies." These discrepancies can be interesting in and of themselves, in that they can reveal, say, tax evasion (see section **??** for an example). But for the purposes of the WID they are more of a nuisance, and we prefer to eliminate them. Second, they can appear because we combine heterogeneous sources, so that various subcomponents of income may have been computed using slightly different methodologies.

Obviously, we stress that these discrepancies must remain reasonable if we want to be able to correct them. In fact, finding large contradictions of accounting identities is the main way we identify plausibility issues in the raw data. We either remove the data or correct it by hand. But for the remaining inconsistencies (representing a few percentage points of GDP at most), we have developed a specific Stata command named `enforce`.[2]

[2] See `https://github.com/thomasblanchet/enforce`. Available on SSC: install by typing `ssc install`

The command is designed to enforce an arbitrary set of accounting identities intelligently, while performing a series of auxiliary checks and adjustments. To understand the command, first assume an identity $a = b + c$. The simplest, naive way of enforcing it is to multiply both b and c by the same constant, $a/(b+c)$. The main problem with that approach is that it only works well with positive variables. If $a = 0$, it will set both b and c to zero, instead of enforcing the more general condition $b = -c$. If $b + c = 0$, then it will not work at all, and more generally if $b + c \ll a$ (because $b < 0$ or $c < 0$), it can easily lead to absurd adjustments.

One way to fix that first issue is to calculate the discrepancy $\varepsilon = a - b - c$, and redistribute it proportionally to the absolute value of b and c. That is, we redefine b as $b + \lambda\varepsilon$ and c as $c + (1-\lambda)\varepsilon$, where $\lambda = |b|/(|b| + |c|)$. If b and c are both positive, this is equivalent to the naive approach, but otherwise it behaves much more reasonably.

But there are still other problems. First, it forces us to define a reference variable (in this case a) that will remain unchanged, which may or may not be desirable. Second, it is not clear how to generalize this adjustment to more complex settings. In practice, we must simultaneously satisfy dozens of accounting identities, with variables present in several of them, so that adjustments must be performed across several dimensions. We formalize that problem as follows. Assume that we have a vector $X = (x_1, \dots, x_n)'$ of variables to be adjusted, and we seek an adjusted vector $Y = (y_1, \dots, y_n)'$ that must satisfy a set of accounting identities. We will minimize:

$$\sum_{i=1}^{n} \frac{(y_i - x_i)^2}{|x_i|}$$

subject to the accounting identities. The convex cost function $(y_i - x_i)^2$ at the numerator ensures that the differences between raw and adjusted variables are as low as possible, and that they are spread equitably across all the variables. The $|x_i|$ at the denominator ensures that adjustments are penalized in proportion to the initial value of the variable. In simple cases, the result is equivalent to the procedure explained above.

Assume that the set of accounting identities can be written as a linear system $AX = B$. The problem can be written in matrix form as:

$$\text{minimize} \qquad \frac{1}{2}X'QX + C'X \qquad \text{subject to} \qquad AX = B$$

where $Q = \text{diag}(1/|x_1|, \dots, 1/|x_n|)$ and $C = (\text{sign}(x_1), \dots, \text{sign}(x_n))'$. This is a standard, qua-

`enforce` in Stata.

dratic programming problem with linear equality constraints and a positive definite matrix Q. Thus, the result is the solution of the linear system:

$$\begin{bmatrix} Q & A' \\ A & 0 \end{bmatrix} \begin{bmatrix} X \\ \lambda \end{bmatrix} = \begin{bmatrix} -C \\ B \end{bmatrix}$$

where λ is a set of Lagrange multipliers determined alongside X. (Variables that are set fixed or equal to zero are removed from the vector X and included in B.) This is the main task performed by the command `enforce`, but it also has several additional features.

First, it takes advantage of accounting identities to fill in any missing value that can technically be calculated from nonmissing variables, even though it was initially absent from the raw data. Second, it pays specific attention to the way constraints are defined to overcome missing value problems. Indeed, assume for example that we have the constraints $a = b + c$, $\alpha = a + x$, $\beta = b + y$, $\gamma = c + z$, and $x = y + z$. Clearly, this implies that $\alpha = \beta + \gamma$. However, if the data were to only contain nonmissing values for α, β and γ, a naive treatment of missing values would lead us to dismiss all the constraints as irrelevant to our data. The command is designed to be aware of the fact that the system of identities implicitly imposes $\alpha = \beta + \gamma$, by performing various singular values decompositions of the system of constraints. Third, the command analyzes the system of identities to find any implausibility (e.g., variable always equal to zero) or incompatibility with the data (in case of fixed variables). Fourth, it provides extensive reporting on the magnitude of the discrepancies and the adjustments.

We use the command `enforce`, to guarantee the consistency of the data both before and after combing them. The resulting database therefore contains no unnecessary missing values, and no statistical discrepancies.

4.1.2.2 Imputation of Consumption of Fixed Capital

CFC is central to our database because it is necessary to compute net national incomes. Its sectorial breakdown is also important because it allows us to properly compare several subcomponents of national income, which are otherwise reported gross or net of CFC depending on the period or the country.

Unfortunately, CFC is not always estimated by national accountants, and sometimes the estimated values must be dismissed because they are implausible. As a result, coverage of CFC is rather poor (around 10–15% of country/years). We have developed a simple imputation method that incorporates three stylized facts about CFC:

- CFC tends to represent a higher fraction of GDP in more developed countries.
- Some countries have structurally high (or low) levels of CFC.
- CFC as a share of GDP is persistent: that is, if CFC is unusually high in year t, it will generally also be high in year $t+1$.

We thus model CFC as a share of GDP as a function of GDP per capita at PPP, using a log-log specification. The model includes a random effect that capture constant country characteristics. Using the index t for the years, and i for the countries, we have:

$$y_{it} = \beta_0 + \beta_1 x_{it} + \beta_2 x_{it}^2 + u_i + \varepsilon_{it}$$

where y_{it} is the logarithm of CFC as a fraction of GDP, x_{it} is the logarithm of GDP per capita at PPP, u_i is the random effect term, and ε_{it} is the error term. The square of x_{it} lets us capture the concavity of the relationship between CFC and GDP per capita. We smooth GDP using the Hodrick-Prescott filter before performing the analysis to avoid capturing short term variations of output, which would make CFC countercyclical. As in any random effect model, we assume:

$$\mathbb{E}[x_{i1}, \dots, x_{iT}] = 0$$

To take into account the persistence of CFC, we model the error term ε_{it} as an AR(1) process:

$$\varepsilon_{it} = \rho \varepsilon_{i,t-1} + \eta_{it}$$

where η_{it} is and i.i.d. white noise. The model can be estimated by generalized least squares using Stata's `xtregar` command. We use the estimated model to impute missing data, taking carefully into account the autocorrelation of the residual. For that, we assume that the white noise is Gaussian: the imputation therefore amounts to a form of kriging (e.g. Chilès and Desassis, 2018), for which we detail the computations below. In order to get a point estimate, all we need is to predict the expected value of the residual at any point in time, but information on the variance remains useful to know the quality of the imputation.

Imputations for the End of the Series Assume that we know the value of ε_T, and we wish to predict the values of $\{\varepsilon_{T+1}, \varepsilon_{T+2}, \dots\}$. By induction, we have $\varepsilon_{T+t} = \rho^t \varepsilon_T + \sum_{k=1}^{t} \rho^{t-k} \eta_{T+k}$. Since ε_{T+t} follows a normal distribution, we only need to calculate its mean and its variance to characterize it. Because ε_T is independent from the future innovations $\{\eta_{T+1}, \eta_{T+2}, \dots, \eta_{T+t}\}$, we simply have:

$$\mathbb{E}[\varepsilon_{T+t}|\varepsilon_T] = \rho^t \qquad \mathrm{Var}(\varepsilon_{T+t}|\varepsilon_T) = \sigma_\eta^2 \frac{1-\rho^{2t}}{1-\rho^2}$$

Imputations for the Beginning of the Series Next, we consider the case where we know the value of ε_0, and we want to know the distribution of $\{\varepsilon_{-1}, \varepsilon_{-2}, \dots\}$ given ε_0. We cannot proceed as simply as we did for the end of the series, because as opposed to future innovations, past innovations $\{\eta_{-1}, \eta_{-2}, \dots\}$ are not independent from ε_0. Using the recurrence relationship of the AR(1) process backward, we get $\varepsilon_{-t} = \frac{1}{\rho^t}\varepsilon_0 - \sum_{k=0}^{t-1} \frac{1}{\rho^{t-k}}\eta_{-k}$. Therefore:

$$\mathbb{E}[\varepsilon_{-t}|\varepsilon_0] = \frac{1}{\rho^t}\varepsilon_0 - \sum_{k=0}^{t-1} \frac{1}{\rho^{t-k}}\mathbb{E}[\eta_{-k}|\varepsilon_0]$$
$$\mathrm{Var}(\varepsilon_{-t}|\varepsilon_0) = \sum_{i=0}^{t-1}\sum_{j=0}^{t-1} \frac{1}{\rho^{2t-i-j}}\mathrm{Cov}(\eta_{-i}, \eta_{-j}|\varepsilon_0)$$

We can write ε_0 as $\varepsilon_0 = \sum_{k=0}^{+\infty} \rho^k \eta_{-k}$. Hence $\mathrm{Cov}(\eta_{-k}, \varepsilon_0) = \rho^k \sigma_\eta^2$. Let $k \in \{0, 1, \dots\}$, and define the Gaussian vector:

$$\begin{bmatrix} \eta_{-k} \\ \varepsilon_0 \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_\eta^2 & \rho^k \sigma_\eta^2 \\ \rho^k \sigma_\eta^2 & \sigma_\eta^2/(1-\rho^2) \end{bmatrix} \right)$$

We can condition the first component of the vector on the second using standard formulas (e.g. Petersen and Pedersen, 2012), which leads to $\mathbb{E}[\eta_{-k}|\varepsilon_0] = \rho^k(1-\rho^2)\varepsilon_0$ and $\mathrm{Var}(\eta_{-t}|\varepsilon_0) = \sigma_\eta^2[1-\rho^{2k}(1-\rho^2)]$. Then, for $i, j \in \{0, 1, \dots\}$ $(i \neq j)$, define another Gaussian vector:

$$\begin{bmatrix} \eta_{-i} \\ \eta_{-j} \\ \varepsilon_0 \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_\eta^2 & 0 & \rho^i \sigma_\eta^2 \\ 0 & \sigma_\eta^2 & \rho^j \sigma_\eta^2 \\ \rho^i \sigma_\eta^2 & \rho^j \sigma_\eta^2 & \sigma_\eta^2/(1-\rho^2) \end{bmatrix} \right)$$

Using the same formulas as before, we get $\mathrm{Cov}(\eta_{-i}, \eta_{-j}|\varepsilon_0) = -\sigma_\eta^2 \rho^{i+j}(1-\rho^2)$. And finally, after simplifications:

$$\mathbb{E}[\varepsilon_{-t}|\varepsilon_0] = \rho^t \varepsilon_0 \qquad \mathrm{Var}(\varepsilon_{-t}|\varepsilon_0) = \sigma_\eta^2 \frac{1-\rho^{2t}}{1-\rho^2}$$

Imputations Between Gaps Finally, consider the case where we know ε_0 and ε_{T+1}, and we want to determine the distribution of $\{\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_T\}$. As usual, we have $\varepsilon_t = \rho^t \varepsilon_0 + \sum_{k=1}^{t} \rho^{t-k} \eta_k$. Hence, the distribution of innovations $\{\eta_1, \eta_2, \ldots, \eta_T\}$ is such that $\sum_{k=1}^{T+1} \rho^{T+1-k} \eta_k = \varepsilon_{T+1} - \rho^{T+1} \varepsilon_0$. We can proceed similarly as in the previous section, except that we condition on both ε_{T+1} and ε_0. We have:

$$\mathbb{E}[\varepsilon_t | \varepsilon_0, \varepsilon_{T+1}] = \rho^t \varepsilon_0 - \sum_{k=1}^{t} \rho^{t-k} \mathbb{E}[\eta_k | \varepsilon_0, \varepsilon_{T+1}]$$
$$\text{Var}(\varepsilon_t | \varepsilon_0, \varepsilon_{T+1}) = \sum_{i=1}^{t} \sum_{j=1}^{t} \rho^{2t-i-j} \text{Cov}(\eta_i, \eta_j | \varepsilon_0, \varepsilon_{T+1})$$

For $k \in \{1, 2, \ldots, t\}$, define the Gaussian vector:

$$\begin{bmatrix} \eta_k \\ \varepsilon_{T+1} - \rho^{T+1} \varepsilon_0 \end{bmatrix} \sim \mathscr{N} \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_\eta^2 & \rho^{T+1-k} \sigma_\eta^2 \\ \rho^{T+1-k} \sigma_\eta^2 & \sigma_\eta^2 \frac{1-\rho^{2(T+1)}}{1-\rho^2} \end{bmatrix} \right)$$

Conditioning the first component of the vector on the second yields:

$$\mathbb{E}[\eta_k | \varepsilon_0, \varepsilon_{T+1}] = \frac{\rho^{T+1-k}(1-\rho^2)}{1-\rho^{2(T+1)}} (\varepsilon_{T+1} - \rho^{T+1} \varepsilon_0)$$
$$\text{Var}[\eta_k | \varepsilon_0, \varepsilon_{T+1}] = \sigma_\eta^2 \left[1 - \frac{\rho^{2(T+1-k)}(1-\rho^2)}{1-\rho^{2(T+1)}} \right]$$

Then, for $i, j \in \{0, 1, \ldots\}$ ($i \neq j$), define another Gaussian vector:

$$\begin{bmatrix} \eta_i \\ \eta_j \\ \varepsilon_{T+1} - \rho^{T+1} \varepsilon_0 \end{bmatrix} \sim \mathscr{N} \left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_\eta^2 & 0 & \rho^{T+1-i} \sigma_\eta^2 \\ 0 & \sigma_\eta^2 & \rho^{T+1-j} \sigma_\eta^2 \\ \rho^{T+1-i} \sigma_\eta^2 & \rho^{T+1-j} \sigma_\eta^2 & \sigma_\eta^2 \frac{1-\rho^{2(T+1)}}{1-\rho^2} \end{bmatrix} \right)$$

From there we get $\text{Cov}(\eta_i, \eta_j | \varepsilon_0, \varepsilon_{T+1}) = -\sigma_\eta^2 \rho^{2(T+1)-i-j} (1-\rho^2)/(1-\rho^{2(T+1)})$. Finally, after simplifications:

$$\mathbb{E}[\varepsilon_{-t} | \varepsilon_0, \varepsilon_{T+1}] = \rho^t \varepsilon_0 + (\varepsilon_{T+1} - \rho^{T+1} \varepsilon_0) \frac{\rho^{T+1-t}(1-\rho^{2t})}{1-\rho^{2(T+1)}}$$
$$\text{Var}(\varepsilon_{-t} | \varepsilon_0, \varepsilon_{T+1}) = \sigma_\eta^2 \frac{(1-\rho^{2t})(1-\rho^{2(T+1-t)})}{(1-\rho^2)(1-\rho^{2(T+1)})}$$

Correction for low CFC values Unlike most components of the System of National Accounts (SNA), the consumption of fixed capital (CFC) is not directly observable. Its estimation is complex and may vary across countries. To address instances where reported or imputed CFC values appear implausibly low, we apply a minimum threshold. Specifically, we set CFC to a minimum of 5% of GDP for years after 1980, and 4% of GDP for the period between 1950 and 1980.

Sectorial Decomposition of CFC When no data is available, we assume that the split of CFC between the different sectors is the same as the median split observed in other parts of the world. Within the household sector, unless we have direct data, we assume that CFC falls onto operating surplus and mixed income in proportion of gross mixed income income and 30% of gross operating surplus.

4.1.2.3 Estimation of Reinvested Earnings on Foreign Portfolio Investment

Our motivation for estimating a flow of "reinvested earnings on foreign portfolio investment" was explained in section 2.1.1.1. Our correction procedure follows the principles set forth by Allen (2019, chap. 4), with a few modifications (namely, we focus on net rather than gross corporate savings, treat investment fund shares differently, and make some imputations to cover the entire world).

We start by estimating, in each country, the share of net corporate savings that is attributable to foreigners. In official statistics, the net saving flow of corporations excludes investment funds and corporations owned by foreigners via foreign direct investment, but otherwise include corporations owned by domestic residents, and corporations owned by foreigners through portfolio investment. International investment position statistics from the IMF use different categories: they distinguish portfolio from foreign direct investment, but they do not separate investment funds.

Using the financial balance sheets of countries from the OECD, define A the total equity liability of the domestic economy, including investment fund shares (F5, S1), and A' the total equity liability of the domestic economy, excluding investment fund shares (F51, S1). Similarly define B (resp. C) the total equity assets (resp. liabilities) of the rest of the world (F5, S2), and B' (resp. C') the total equity assets (resp. liabilities) of the rest of the world, excluding investment fund shares (F51, S2). Also define the ratios $\beta = B'/B$ and $\gamma = C'/C$.

Then, we turn to international investment position statistics, using official data from the IMF, and the data from Lane and Milesi-Ferretti (2018). This data offers a wider coverage than country balance sheets, but they do not separate investment fund share. For each country, we extract the value E of portfolio liabilities from this source, and multiply it by the ratio β calculated above. We carry this ratio forward (or backward) in time to overcome the more limited time coverage, and assume that it is equal to one when it is not observed at all. We also extract D, the net FDI position.

Then we estimate the share of net corporate savings that is attributable to foreigners in each country as $\beta E/(A'+D)$. This ratio cannot be calculated unless we observe A', for which we have a much sparser coverage than βE or D. However, most of the variation in $\beta E/(A'+D)$ is driven by the numerator rather than the denominator, so that the correlation between $\beta E/(A'+D)$ and βE is around 85%. Therefore, we impute and extrapolate the value of $\beta E/(A'+D)$ based on βE when we can only observe βE. Still, for most countries, the data does not start until the 1990s. However, the rise of foreign portfolio liabilities is a rather recent phenomenon. So, unless we observe otherwise, we assume that $\beta E/(A'+D)$ was zero in 1970 and rose linearly until its first observed value.

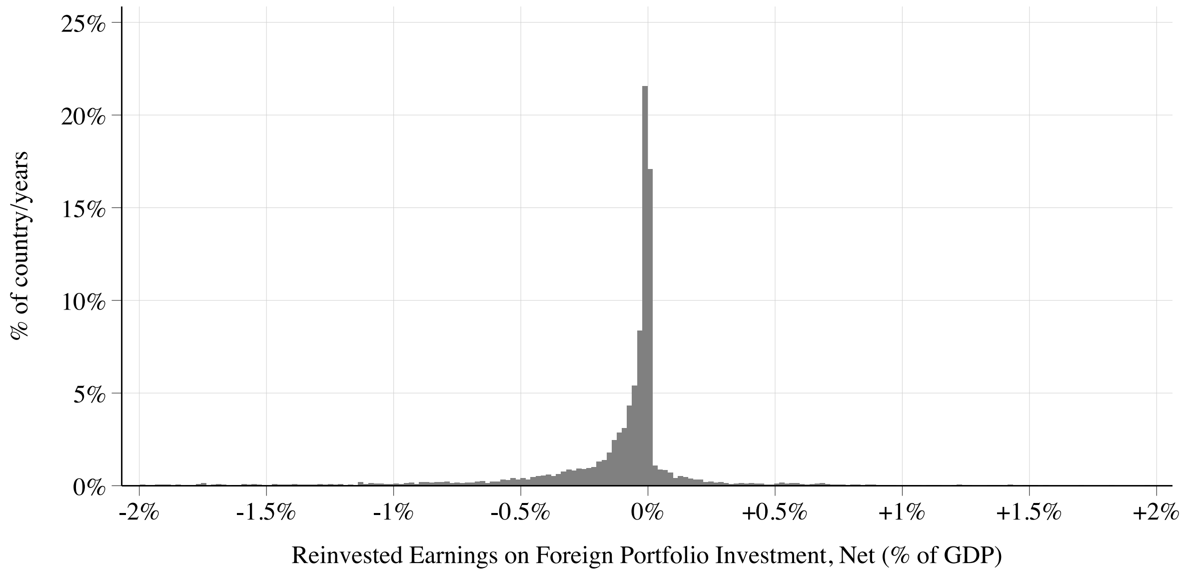
	France	**Germany**	**United States**
Equity liability of the domestic economy, excluding investment fund shares: A'	7,092	3786	51,111
Equity assets owned by the rest of the world, including investment fund shares: B	2,183	2,455	11,940
Equity assets owned by the rest of the world, excluding investment fund shares: B'	1,912	1,911	11,232
Ratio: $\beta = B'/B$	88%	78%	94%
Net FDI position: D	539	589	395
Portfolio liabilities: E	848	731	6,219
Share of corporate savings attributable to foreigners: $\beta E/(A'+D)$	10%	13%	11%

Source: WIL estimates using the IMF BOPS, Lane and Milesi-Ferretti (2018), OECD Financial balance sheets.
Note: monetary amounts in billion current USD.

Table 4.1: Estimates of the Foreign Share of Retained Earnings, 2015

The amount of corporate savings that we attribute to foreigners is equal to net corporate savings multiplied by $\beta E/(A'+D)$. In countries where corporate savings are not observed, we rely on regional averages (as a fraction of GDP). We illustrate this estimate in table 4.1. In the three countries given as examples (France, Germany and the United States), the share of retained earnings attributable to foreigners is around 10%.

We then reallocate these foreign corporate savings to their country of destination using bilateral data on foreign investment positions from the IMF CPIS. This data gives us the amount of foreign portfolio equity in country X that belongs to country Y. First, we multiply that value by the ratio γ of country Y as calculated above to correct for the inclusion of investment fund shares. Then, we estimate that if, say, 10% of the adjusted equity liability of country X belongs to Y, then 10% of the net corporate savings of country belong to country Y.



Source: WIL estimates using the IMF BOPS, IMF CPIS, Lane and Milesi-Ferretti (2018), OECD Financial balance sheets. *Methodology:* see main text. *Note:* The data covers all countries from 1970 to 2019. Bins of equal size of 0.02% of GDP each.

Figure 4.1: Distribution of Estimated Net Foreign Retained Earnings

There are many strong assumptions behind these estimates, and we stress that they will be subject to revisions should better data become available. Most importantly, the methodology relies on the assumptions that foreign-owned firms are similar to domestic ones, and that foreign-owned firms are all similar to one another. In general, the adjustment is quite limited: 90% of country/year estimates (since 1970) are comprised between -1.1% and $+0.6\%$ of GDP, with a majority being close to zero (see Figure 4.1). But there is a small number of large positive outliers: countries with the largest adjustment are the Cayman Islands (up to 360% of GDP), the Netherlands Antilles (up to 160%), Mauritius (up to 120%), Bermuda (up to 75%), Luxembourg (up to 70%), Sint Marteen (up to 8%) and Curaçao (up to 8%). They correspond to small tax havens, and these values are due to their large claims over assets of large economies in the CPIS data (e.g., the Cayman Islands and the United States, or Mauritius and India). There are many reasons why the similarity assumption could be invalid in these cases. To avoid excessive and implausible corrections, we arbitrarily divide the inflow of foreign retained earnings by ten in these rare cases. This is a purely *ad hoc* adjustment which we hope can be improved in the future.

4.2 Imputation of Factor Shares and Sectoral Decomposition

The WID provides estimates for the sectoral decomposition of GDP (or NDP) and factor shares for all 216 WID countries since 1980, and since 1900 for 57 core territories. However many countries do not publish the full sectoral decomposition of GDP in their official statistics, especially before 1970-1980. We briefly explain our imputation procedure for such cases below.[3] We stress that we still face many data limitations when comparing institutional sectors and factor shares over time and across countries. The series are not meant to be final. They will be revised and updated as new and hopefully more comparable country estimates become available.

This sectoral decomposition follows the "generation of income account" in the sequence of national accounts and shows in which sector primary income *originates*. This differs from the perspective of the "allocation of income account" used in section 2.1.2, which shows the sectors destined to *receive* such incomes. We can see this difference in the household sector. While all wages received are part of the primary income of the household sector, they are part of the value added of the corporate sector if paid by corporations, part of the government sector if paid by the government or part of the household sector if paid by households.

Imputation of Sectoral Decomposition since 1980 We apply the following steps to complete the sectoral decompositions for countries and years with missing data. First, if countries do report some, but not all items of the sectoral decomposition, we use the values of Bachas, M. H. Fisher-Post, et al. (2022), who obtain the missing national accounts items for 150 countries from 1965 based on a regression on NNI per capita. To get the full sectoral decomposition, the compensation of employees paid by households is missing. We calculate the total compensation of employees paid by the private sector by deducting compensation of employees paid by the government and net foreign labor income from the total compensation of employees received. We split this remaining privately paid compensation of employees based on the last year with available data, or, if this is never observed, according to the relative size of mixed income and corporate profits as a proxy. In any case the compensation of employees paid by households is usually small (1% to 5% of GDP) and this split does not affect the capital share of the total economy. For countries not covered by Bachas, M. H. Fisher-Post, et al. (2022) or those where

[3] A more detailed explanation is provided in Dietrich et al. (2025)

all years are based purely on imputations, we use regional averages of observations post 1950 based on five GDP per capita bins to assign sectoral value added. If available, we keep the share of wages paid by the corporate or household sector constant for each country. If they are never observed in a country, we use the regional average shares.

Historical imputation for 57 core territories since 1900 For the first half of the 20th century data on GDP by institutional sector is very rare. However, in all countries with data we observe that increasing GDP per capita comes together with a decline in the value added by the household sector and the rise of corporations. We use this regularity to estimate sectoral shares for all core territories with missing data from 1900 to 1979. Based on historical series for France, USA, United Kingdom and Sweden, we provide estimates on sectoral GDP by three GDP per capita bins. These values have the following trends that we observe in historical data: a decrease in the production by the household sector, an increase in the corporate sector, as well as increase in the government sector and an increase in CFC. We use these values to assign a sectoral decomposition for all 57 core territories with missing data in 1900 and 1930 based on their GDP per capita and interpolate linearly. To estimate historical factor shares, we further assume a constant gross profit share in the corporate sector of 40% and a constant net profit share in the corporate sector of 35% of corporate GVA for countries in Europe, North America & Oceania, East Asia or Russia & Central Asia. For countries in South & South-East Asia, Subsaharan Africa, Latin America or MENA we assign a gross corporate sector gross profit share in the corporate sector of 55% and a constant net corporate profit share of 50% of corporate GVA. We do so to account for the very large regional differences in capital shares within the corporate sector in recent decades. While this strategy is clearly imperfect, we prefer these simple and transparent assumptions to more complex and "black-box" procedures. These estimates will be revised and improved as more country studies become available.

4.3 Wealth Aggregates

The macroeconomic wealth side of the WID contains the decomposition of national wealth into resident sectors (i.e., households, non-profit institutions serving households, corporations and the government) and the foreign assets and liabilities *vis-à-vis* the rest of the world. This section briefly explains our estimates of aggregate wealth at WID, a more detailed description is provided in Bauluz, Brassac, et al. (2024).

Data & Methodology. As a general rule, we rely on official national accounts, academic studies or secondary sources to reconstruct sectoral balance sheets across countries worldwide. We start from authors' works who assemble full balance sheets, mostly built upon official statistics. Since SNA 2008-compatible official balance sheets often start in the 1990s, researchers extend these official series backwards by using a combination of historical balance sheets from official sources (i.e., based on SNA 1993 concepts and pre-SNA standards) and estimates using historical records. The general procedure followed to estimate aggregate wealth series is explained by Piketty and Zucman (2014), who produced long-run series of aggregate wealth for eight advanced economies: Australia, Canada, France, Germany, Japan, Italy, the United Kingdom and the United States. The adaptation of Piketty and Zucman's (2014) series to the most recent national accounts system (SNA 2008) and the latest WID guidelines has been tackled by Bauluz (2017). This pioneering work has been pursued by various authors – Albers, Bartels, and Schularick (2025), Artola Blanco, Bauluz, and Martínez-Toledano (2021), Baselgia and Martínez (2023), Carranza, De Rosa, and Flores (2025), Chatterjee, Czajka, and Gethin (2020), Daly and Morgan (2021), Kumar (2019), Mo et al. (2025), Novokmet, Piketty, and Zucman (2018), Toussaint et al. (2025), and Waldenström (2017) – thanks to whom we are now able to provide long-run series of national wealth for most advanced economies and for a few developing countries (e.g.: Brazil, China, India, South Africa, Uruguay), dating back to the early 19th century for some countries. That selection and the time span largely depend on the availability of official estimates of wealth aggregates, and the capacity of researchers to systematically exploit historical sources to generate long-run series of wealth aggregates.

When full balance sheet data are not available, we collect information on specific subcomponents. This is often the case for financial assets and liabilities. We also rely on official financial balance sheets provided by National Central Banks, National Statistical Offices, Eurostat, the OECD, or the IMF, which typically offer broader coverage than non-financial balance sheets. When the latter do not exist, we also build on existing databases like the OECD Pension Markets in Focus for household pension assets, the IMF Global Debt Database for public debt or Nievas and Sodano (2024) for foreign wealth.

Imputation Method. For the period 1980-onwards, whenever we do not observe these balance sheet data or data on specific wealth components, we provide wealth estimates for the whole world, broken down by sector and main asset categories, through a simple and transparent imputation procedure. More specifically, we cover the 216 countries defined by the WID and estimate private, corporate and public: housing assets, business assets, financial assets and liabilities. We go further when it comes to private wealth by distinguishing agricultural land from other domestic capital and within private financial assets: fixed-income assets, equities and pension assets. Paired with our foreign wealth data drawn from Nievas and Sodano (2024), we are in turn able to compute market & book-value national wealth estimates as well.

We also provide estimates of the breakdown of private wealth between households and non-profit institutions serving households (NPISH) for the entire world since 1980. When household wealth is not directly observed for a given country, we estimate it by multiplying total private wealth by the average household-to-private wealth ratio observed in countries where both measures are available in the raw (non-imputed) data. This ratio is approximately 0.95, implying that around 5% of private wealth is owned by non-profit institutions serving households. Consequently, we also produce estimates of NPISH wealth, calculated as the difference between private net wealth and household net wealth.

Our imputation procedure to estimate missing data relies on a simple OLS regression framework, where a missing dependent variable[4] is predicted based on a set of covariates (e.g. population, net national income, top 10% income share, life expectancy at birth, urban population, population density, capital shares, foreign wealth, private and public capital stock; mostly drawn from the WID or the World Bank). Thus, the model requires data on both the dependent and independent variables across countries to generate reliable predictions. In this respect, our approach follows Davies, Lluberas, and Shorrocks (2017) and Davies, Sandström, et al. (2011), who use a similar method to estimate household financial assets, non-financial assets, and debt since 2001. The method is transparent and provides fairly good approximations when the predicted data is compared with the observed data.

The coverage of these variables in the raw data varies across time and world regions. While data coverage is higher among high-income countries, we also generally have observations for large developing economies, including China, India, South Africa, Russia, Brazil, and Mexico. On average, we have data for approximately 20% of all country-year pairs, which account for around two-thirds of global national income (because larger and richer countries are better covered).

While the OLS framework described above serves as the foundation for most imputations, we opt for a more straightforward method for agricultural land and government non-financial assets. We do so because we count with data very closely related to these two variables. For government non-financial assets, estimates of public fixed capital (produced non-financial assets) are available from the IMF for a very large sample of countries since the 1970s. Since total non-financial assets is the sum of fixed (produced) and non-produced assets, we rescale the

[4] See list above: housing assets, business assets, financial assets and liabilities of the private, corporate and public sectors. But also for the private sector: agricultural land, other domestic capital, fixed-income assets, equities and pension assets.

IMF data using the average ratio of total non-financial assets to fixed assets in countries where both variables are available, which is approximately 1.5. For private agricultural land, we make use of FAO statistics on Gross Value Added in agriculture to impute agricultural land, which are available on an annual basis for most countries in the world and go back to the 1970s. We predict the (log) ratio of agricultural land to GDP using the (log) ratio of GVA in agricultural land to GDP. We impute national agricultural land by rescaling private agricultural land with the average ratio in countries where both are observed, which is approximately 1.41.

When data unavailability prevents us from making OLS predictions, we extrapolate values both forward and backward using trends from predicted values or regional averages. When data is fully imputed, we extend estimates using regional trends. Additionally, we assign regional averages to fill remaining gaps.

We refer readers interested in the precise sources for each country or details about the imputation method — like how we deal with former Soviet States in the 1980s or outliers at the beginning and end of our imputation method -– to Bauluz, Brassac, et al. (2024).

Foreign Wealth. With regards to foreign wealth, we aim to enforce two major accounting rules, while providing world coverage since 1980: 1) Global net foreign wealth should add up to to zero. 2) Net domestic financial assets should be equal to net foreign assets at the country-level. To achieve these objectives, we rely on the foreign wealth series from Nievas and Sodano (2024) -– built upon net international investment position (IIP) data -– which encompass all 216 WID countries since 1970 and reach net zero foreign wealth at the world-level. To reach the later, we must tackle an issue of practical order. Conceptually, net domestic financial assets are equal to net foreign assets at the country-level[5]. While it is usually the case in the data, there are, however, some instances in which net domestic financial assets and net foreign assets do not coincide, in particular when net foreign assets series are not derived from official financial balance sheet data but from the balance of payments (see Piketty and Zucman, 2014, Appendix, section A.4.4, pp. 31-32). We classify these discrepancies into three types. The first one involves inconsistencies within official financial balance sheets between the domestic and foreign sectors. The second type refers to inconsistencies between net foreign assets in official financial balance sheets and IIP in the balance of payments – due to monetary gold or the accumulation of "error and omissions". The third arises because part or all the components of the net domestic financial assets are obtained from OLS regressions, leading to differences between the predicted net

[5]In a closed economy, a sector's financial asset is another sector's liability, so financial assets cancel out liabilities – with the exception of monetary gold. In an open economy, the difference between the two represents the external position.

domestic financial assets and the observed IIP from Nievas and Sodano (2024). To address these discrepancies, we enforce at the country-level that net domestic financial assets must equal net foreign assets. To do so, we compute the discrepancy between the two, allocate it to corporate financial wealth and accordingly rescale variables derived from it: corporate book value, residual corporate wealth, Tobin's Q, and national book-value wealth. It must however be noted that when these discrepancies do arise, they are generally small or negligible. We refer readers specifically interested in the matter to Bauluz, Brassac, et al. (2024, section 2.8).

4.4 The Balance of Payments

4.4.1 Foreign Capital Income

The primary source for data on foreign capital income is the IMF Balance of Payments (BOP), and in situations where IMF data is not accessible, alternative sources like the United Nations or OECD statistics are used (for a detailed coverage see Nievas and Sodano (2024). Foreign capital income encompasses diverse components, including portfolio and other income received and paid, and reinvested earnings on portfolio investment.

If foreign capital income is not reported for a certain year but an aggregate is reported (e.g.: foreign income), then we use the foreign capital income-to-foreign income ratio of the closest year to fill in the missing value. If foreign capital income received or paid is available but the country does not report its decomposition (FDI or portfolio), then we assume each asset class capital income is proportional to the share of the asset class on aggregate wealth.

For missing values, predictions are made based on asset class stock, GDP in USD, exchange rates, and inflation rates. Return rates predictions are made separately for each asset class since FDI is assumed to be more profitable than portfolio. An Ordinary Least Squares (OLS) regression model is used, including country-specific fixed effects to account for time-invariant characteristics of each economy, as well as region-year fixed effects to capture unobserved shocks affecting the region uniformly. Specifically:

$$i^B_{\rho,c,t} = \beta_0 + \beta_1 \frac{\text{wealth}^B_{\rho,c,t}}{\text{GDP}_{c,t}} + \beta_2\, e_{c,t} + \beta_3\, \pi_{c,t} + \alpha_c + \gamma_{r,t} + \varepsilon_{c,t}\,. \tag{4.1}$$

Where i refers to the return rate, B to asset or liability, ρ to the asset class (FDI or portfolio), c to the country, t to the year, e to the nominal exchange rate with respect to US dollars, π to the inflation rate and α, γ and ε to the country fixed effects, region-year fixed effects and error term, respectively. Whenever data is still missing, we impute the value based on the regional average.

4.4.2 Foreign Wealth

The data on foreign wealth is sourced from "The External Wealth of Nations" (Lane and Milesi-Ferretti, 2018), which provides a standard breakdown of external assets and liabilities based on the Balance of Payments (BOP) Statistics Manual 6. External financial assets and liabilities encompass various components, such as foreign direct investment, portfolio equity, portfolio debt, other invest- ment, and financial derivatives. Notably, foreign exchange reserves are included as financial assets, while gold holdings are excluded. In cases where data coverage is incomplete, countries are assumed to follow the regional trend. Only six countries have been completely imputed using a regional average

4.4.3 Corrections

To guarantee that global net foreign capital income and net foreign wealth together sum to zero, we distribute any residual imbalance proportionally across economies. For a detailed comparison of this adjustment procedure and its advantages over alternative methods, see Nievas and Piketty (2024).

Beyond these aggregate adjustments, we also correct how retained earnings (or undistributed profits) on portfolio investment are recorded in national accounts. Retained earnings are the profits that firms keep after paying suppliers, employees, shareholders, and taxes, they are counted as part of national income. Under the current SNA framework, however, if a company in country A is owned by portfolio investors in country B (who lack direct control over management), all its retained earnings are attributed to country A, even though the ultimate beneficiaries are abroad.

We reallocate those profits to reflect true ownership: We estimate (a) the flow of foreign retained earnings that accrue to residents and (b) the flow of domestic retained earnings that accrue to foreigners, and then apply the net of these two to each economy. We carry out this correction for all 216 countries—excluding tax havens—and again verify that the global aggregate sums to zero.

Next, we complete the coverage of IMF BoP trade in goods by relying on well-recorded bilateral statistics (Conte, Cotterlaz, Mayer, et al., 2022). Following common practice in the trade literature, we "mirror" exports to match imports. That is, we assume imports recorded by country A from country B equal exports recorded by B to A, so that, year by year, global exports of goods exactly equal global imports. For the remaining current-account components (services, compensation of employees, other primary and secondary income) and the capital account, the appropriate global balancing adjustment is less clear; accordingly, we simply scale credits and debits proportionally to meet the average of total credits and total debits, whenever the net global balance deviates from zero.

4.5 Historical Macroeconomic Accounts

The WID publishes a harmonized set of annual national-accounts series spanning the period 1800–2023. Drawing on both existing WID data and newly collected historical sources (Nievas and Piketty, 2025a,b), these series cover a stable universe of 57 core territories. Presenting nominal and real GDP, price indices, exchange rates, trade balances, balance-of-payments aggregates, foreign income and foreign wealth, the series provide an unprecedentedly rich portrait of global macroeconomic trends over more than two centuries.

Region (Count)	Countries / Territories
East Asia (5)	China, Japan, South Korea, Taiwan Other EASA (OB)
Europe (11)	Britain, Denmark, France, Germany, Italy, Netherlands, Norway, Spain, Sweden, Other W.EUR (OC), Other E.EUR (QM)
Latin America (6)	Argentina, Brasil, Chile, Colombia Mexico, Other LATAM (OD)
Middle East / North Africa (8)	Algeria, Egypt, Iran, Morocco, Saudi Arabia, Turkey, UAE Other MENA (OE)
North America / Oceania (5)	USA, Canada, Australia, New Zealand Other NAOC (OH)
Russia / Central Asia (2)	Russia Other RUCA (OA)
South / South-East Asia (9)	Bangladesh, India, Indonesia, Myanmar, Pakistan, Philippines, Thailand, Vietnam Other SSEA (OJ)
Sub-Saharan Africa (11)	DR Congo, Ethiopia, Kenya, Ivory Coast, Mali, Niger, Nigeria, Rwanda, Sudan, South Africa Other SSAF (OI)

Table 4.2: Geographical Coverage in Historical Estimates

Geographical Coverage For historical macroeconomic accounts, our definition of the world comprises 57 fixed territories, as opposed to the 216 countries represented in contemporaneous estimates. These territories include 48 countries and 9 residual regions, accounting for roughly 85–90 percent of world population and GDP at both market-exchange-rate and purchasing-power-parity terms throughout the entire 1800–2023 interval. For the remaining 10-15% contained in the residual regions, available data sources are not sufficiently detailed so as to include long run series for all underlying countries – except for the 1970-2023 period, where we provide annual series for all 216 countries

Variable Coverage and Conventions Each country–year observation reports nominal and real GDP in local currency (LCU) and U.S. dollars, price indices, market and purchasing-power-parity (PPP) exchange rates, and derived real exchange-rate series. Consistent with WID practice, LCU series use the latest currency unit (e.g. pre-1999 figures for France are expressed in euros), and we introduce a novel "LCU–USD change" indicator to allow users to reconstruct historical-currency values. For residual and world regions we adopt the U.S. dollar as a conventional LCU, endowing these aggregates with their own price and exchange-rate indices. Where early market-exchange-rate data are missing or monetary regimes are complex, we impute a USD peg prior to the availability of reliable quotes, and we transparently flag these assumptions for future updates.

Balancing and Consistency Beyond GDP and prices, the WID series encompass annual exports, imports, and trade balances for goods (disaggregated into primary commodities and manufactured products) and services, as well as inflows and outflows of foreign income, transfers, and net foreign wealth, assets and liabilities, consumption of fixed capital, net domestic product and net national income. All inter-country flows are internally balanced so that global aggregates sum exactly to zero. This rigorous framework ensures that users can analyze long-run macroeconomic inequalities—both within and between countries, with confidence in the underlying national-account consistency.

Key Points

- The WID provides numerous macroeconomic series on income, wealth, population, and prices. These series are obtained by systematically combining all the existing sources under a consistent framework.
- Some key components (CFC, income from tax havens, reinvested earnings on foreign portfolio investment) have to be estimated or imputed based on limited data, at least in certain countries. These estimates are by definition more fragile and more subject to revisions than the rest.
- The data and code that was used to generate this data (at any point in time) is fully archived and freely available online at `https://github.com/widworld/wid-world`.

Chapter 5

Income Distribution Series

This chapter deals with the reconciliation of micro and macro data in the case of income flows. The underlying income concepts were explained in chapter 2, and chapter 6 will deal with wealth stocks. As with the previous chapter, we only explain the general principles, referring readers to methodological and country-specific papers for more details.

We assume in this chapter that we have access to comprehensive income micro files, which include annual information on individual flows of both labor and capital incomes of the entire population, and that we can use these files to compute distributional estimates that are consistent with the totals from national accounts. (See chapter 7 for situations with more limited data.)

Depending on the country and the size of its informal economy, either tax or survey statistics will be dominant, while the other data set will play a supporting role. For instance, in countries with a small informal economy, and where high-quality tax microdata is available, the tax data should be used as the "main" source. In those cases, income surveys are used to make minor adjustments in order to account for non-filers and certain tax-exempt incomes. We cover this in section 5.1.

On the contrary, in most emerging countries — such as African, Asian or Latin American countries — income surveys will play the primary role, while register data is used to correct the top of the income distribution. In most of these cases, the size of the informal economy — which is not captured by administrative records by definition — is such that tax data cannot be considered as representative for the whole population, especially for middle and lower incomes. In addition, for developed countries in earlier time periods, access to tax micro-files is not an option. Instead, less detailed income tax tabulations, which generally cover a small share of the adult population, are used. In section 5.2, we will discuss the methods that can be used in the case of countries and time periods where tax microdata is not the primary source.

Once the microdata is ready, the basic imputation and estimation methods that we use to produce DINA series are relatively straightforward. That is, we scale fiscal income flows up to national-accounts-based income concepts and impute missing income flows in order to arrive at our benchmark concepts (i.e., pretax factor income, pretax post-replacement income, post-tax disposable income and post-tax national income as defined in chapter 2). We describe this process in sections 5.3 to 5.4. More details are provided in the country-specific papers (see in particular Piketty, Saez, and Zucman (2018) for the United States; Garbinti, Goupille-Lebret, and Piketty (2018) and Bozio et al. (2020) for France; Bach, Bartels, and Neef (2020) for Germany; and De Rosa, Flores, and Morgan (2024) for Latin America).

5.1 Preparing Tax Microdata

In some cases researchers will have access to comprehensive administrative microdata covering a majority of the population. This is the case in a number of developed countries in recent decades. For example, in the United States, we have access to high-quality tax micro-files covering almost the entire population since 1962 (see Piketty, Saez, and Zucman, 2018). In France, we have access to similar microfiles since 1970 (see Bozio et al., 2020; Garbinti, Goupille-Lebret, and Piketty, 2018). In Germany, we have access to triennially issued tax micro files since 1992 (see Bach, Bartels, and Neef, 2020).

But even in those countries, the income tax micro data covers only a subset of the population. Therefore, we have to impute observations for non-filers from representative survey data. Because the aim is to build a dataset representing the entire adult population, the imputation must take into account both the income and demographic characteristics of the adult population.

Income tax data usually reports incomes above a certain income threshold. For example, in Germany, only income tax filers above the basic annual tax allowance are included in the tax data. Thus, the individuals not reaching this income threshold should be imputed. To produce a representative dataset, we recommend using population statistics to impute observations along characteristics such as age, marital status, gender and possibly region of residence. This is important as certain population groups may not be represented equally in income tax statistics. Typical examples are pensioners and young persons, who may be underrepresented due to small taxable incomes. Alternatively, some authors add information on the share of non-filers in the population and their average income from survey data to the analysis.

Further blind spots in the coverage of incomes might emerge due to income specific allowances, such as an annual saver's allowance. Capital incomes under the saver's allowance usually do not have to be reported. Additional imputations might also be necessary for fully tax-exempt items (e.g., imputed rents of owner-occupiers, or certain tax-exempt savings accounts). Several countries have introduced a dual income tax in recent decades (Sweden in 1991, Germany in 2009, or France in 2018 among others) excluding capital incomes from the comprehensive progressive income tax and taxing them at a flat tax withheld at the source. The dual system results in missing information in tax data which must be extrapolated from former years and/or imputed from other sources such as surveys.

In some countries, employee's tax obligations can be settled by their employer. Thus, individuals do not have to submit tax returns unless they also earn income from other sources or want to claim allowances. These incomes do not enter or only enter partially the income tax data. In the case of Croatia, Novokmet and Kump (2018) combine data from the tax withheld at source with data from individual tax returns. In the case of Germany, Bach, Bartels, and Neef (2020) impute missing employee observations from survey data. More details on imputation procedures are provided in the country-specific papers.

5.2 Preparing Survey Microdata

In many countries, both developed and less developed, direct access to rich administrative microdata is quite rare. Instead, it is more common to have tabulations of fiscal income, containing information on a minority of income recipients by income bracket, alongside survey microdata that is more or less representative of the full population. In these cases, rather than incorporating survey information into the tax data, we incorporate tax information into the survey data.

The motive for doing so is that surveys tend to underrepresent high incomes as compared to what appears in register data. This is even more true when the survey has not been matched to administrative records (e.g., social security data on wages). The following sections explains how to proceed. It follows the principles set forth by Blanchet, Flores, and Morgan (2019). To apply the methodology, we provide a Stata command named `bfmcorr` that automatizes most of the steps described below.[1]

[1] See `https://github.com/thomasblanchet/bfmcorr`. Available on SSC: install by typing `ssc install`

5.2.1 Interpolation of Tax Tabulations

The tax tabulations that are provided by tax administrations use arbitrary brackets to describe the distribution. Before we can use this data, we need to recover a complete distribution of taxable income, at least for the part of the distribution covered by the tax. That is, we must "interpolate" the distribution to turn the limited set of brackets into a complete set of g-percentiles. For that, we recommend using the "generalized Pareto interpolation" method developed by Blanchet, Fournier, and Piketty (2017).

This interpolation method, contrary to other available methods (such as the standard Pareto interpolation with a constant coefficient, the log-linear interpolation, or the mean-split histogram) estimates an income distribution without relying on restrictive parametric assumptions. It estimates a varying curve of Pareto coefficients, so that the distribution is allowed to have a flexible form, which produces smoother and more precise estimates of the distribution (see Blanchet, Fournier, and Piketty (2017) for comparisons).

We provide two tools for the researchers to apply the method: a R package named `gpinter` (see https://github.com/thomasblanchet/gpinter) and an online interface based on the package, available at https://wid.world/gpinter/ which can be used without any knowledge of R. We refer to the documentation of these packages for the exact format of the data to be given in input. Note that the package and the online interface include extensions of the interpolation method of Blanchet, Fournier, and Piketty (2017) for cases in which only the threshold (but not the average) of each bracket is available in the tax data, or cases in which only the averages (but not the thresholds) are available. We stress that when only thresholds are available, the interpolation becomes a lot less precise, unless these thresholds cover the very top of the distribution.

5.2.1.1 Combining Survey and Tax Data

Once the tax tabulations have been expanded, we need to know whose income to correct and in which part of the distribution, and for that we must compare the distribution of income in surveys and tax data. To do so, the concepts and units of observation in both datasets must first be as close to each other as possible. This means constructing a variable for "taxable income" in the survey data, i.e., income sources that are liable to be declared on income tax returns. In general this tends to be close to "pretax post-replacement income," i.e., income received by individuals before personal income taxes, deductions, and employees and self-employed social

`bfmcorr` in Stata.

contributions (but after employers' contributions), but after accounting for social insurance benefits (unemployment and pension benefits). We call this variable survey income but, in fact, it should correspond to the taxable income of the population represented in the sample. The remaining differences between the distribution of survey income and taxable income should be due to nonresponse or misreporting, not conceptual discrepancies.

In many cases, survey income is reported net of personal income taxes and social contributions, without any information on the taxes or contributions paid in the survey questionnaire to be able to manipulate the survey concept of income. This can create a discrepancy with the fiscal income concept in tax data. There are two ways around this complication. One is to impute direct taxes and contributions to the survey income using information on rate structures and existing legislation, before relating this information to fiscal income from tax data (which would usually have to be done anyway, see section 5.3). The other is to use available information on tax liability (e.g., tax owed/paid) in the tax tabulations to compute net-of-tax bracket values.[2] Additional information on social contributions (linked to fiscal deductions) can also be used to deduct contributions from total income per bracket. The adjusted tax data income can then be related to survey income. The choice of procedure will depend on the data structure per country.

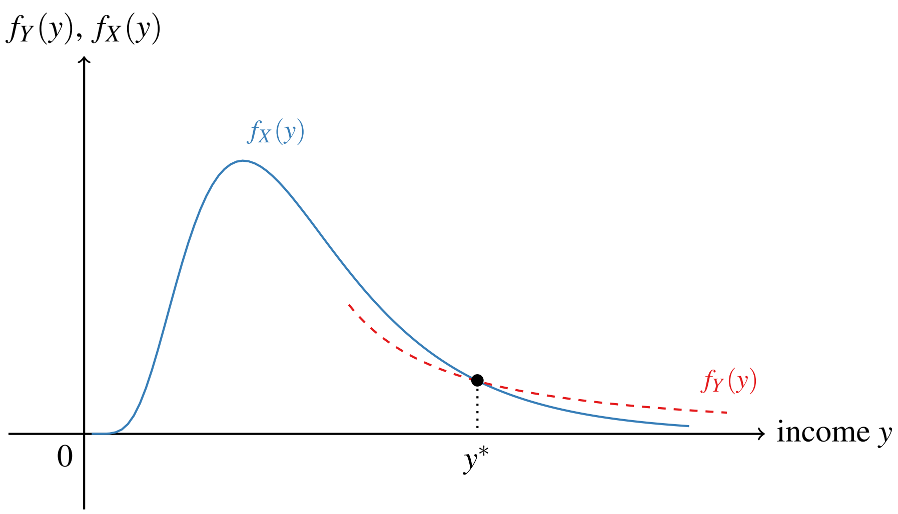


Figure 5.1: Hypothetical Income Distributions in Surveys and Tax Data

When comparing incomes in the interpolated tax distribution to those in the survey distribution, the problem is usually the one presented in figure 5.1. The extent of the rich's underrepresentation in the survey can be viewed by the different densities covered by the two lines. Above a certain

[2]Not that this approach means assuming away potential re-ranking effects from taxes and social contributions on the income distribution.

point, tax data incomes will be both more numerous (dashed line) than incomes in the survey (solid line). Below this point the survey represents more income earners than what it should. The question, then, is how to correct the survey distribution so that it represents the income distribution coming from tax data. We turn to this in the subsequent sections.

5.2.1.2 Overall Method

The method of Blanchet, Flores, and Morgan (2019) is a consistent approach, rooted in calibration theory, for incorporation tax information into the survey. It has three stages: the choice of the "merging point", the reweighting of survey observations above and below this point, and the expansion of the survey's support by including the highest incomes from tax data.

Firstly, the choice of the merging point between the survey and tax distributions requires a comparison of the number of observations within each percentile. The input tax data for this exercise should take the form of the tabulation produced from the interpolation described in section 5.2.1. This should indicate (at least) bracket percentile, bracket thresholds and bracket averages in the form described by table 5.1.

p	thr	bracketavg
0	0	190
0.01	378	568
0.02	757	947
...	...	...
0.99997	10,597,050	12,059,701
0.99998	13,889,397	17,180,048
0.99999	22,006,175	64,444,108

Table 5.1: Example of Tabulation Used to Correct Survey Data

The column p should always express percentiles of the same total population as the survey whose distribution is to be corrected. The method automatically selects a "merging point" between the two distributions by comparing the interpolated distribution of incomes from the tax data above a percentile p^* to the distribution of survey income.

This point p^* should correspond to the beginning of the "trustable span" of the tax data — the point beyond which the distribution is can be considered reliable. This can be defined as the share of the population covered in the tax data or the share of taxpayers in the population.[3]

[3]In practice, taxpayers typically constitute a subset of the population that declares income to the tax authorities.

We stress that the aim is never to use the full distribution from tax data above p^* to correct the survey. This is because the beginning of its trustable span is not always sharp — the reliability of the tax data increases with income but is not always well defined. Therefore it is more prudent to restrict its use to the minimum that is necessary. Moreover, once we are past the point where there is clear evidence of a bias, we prefer to avoid distorting the survey in unnecessary ways. Therefore, we select a merging point that is usually above p^*.[4]

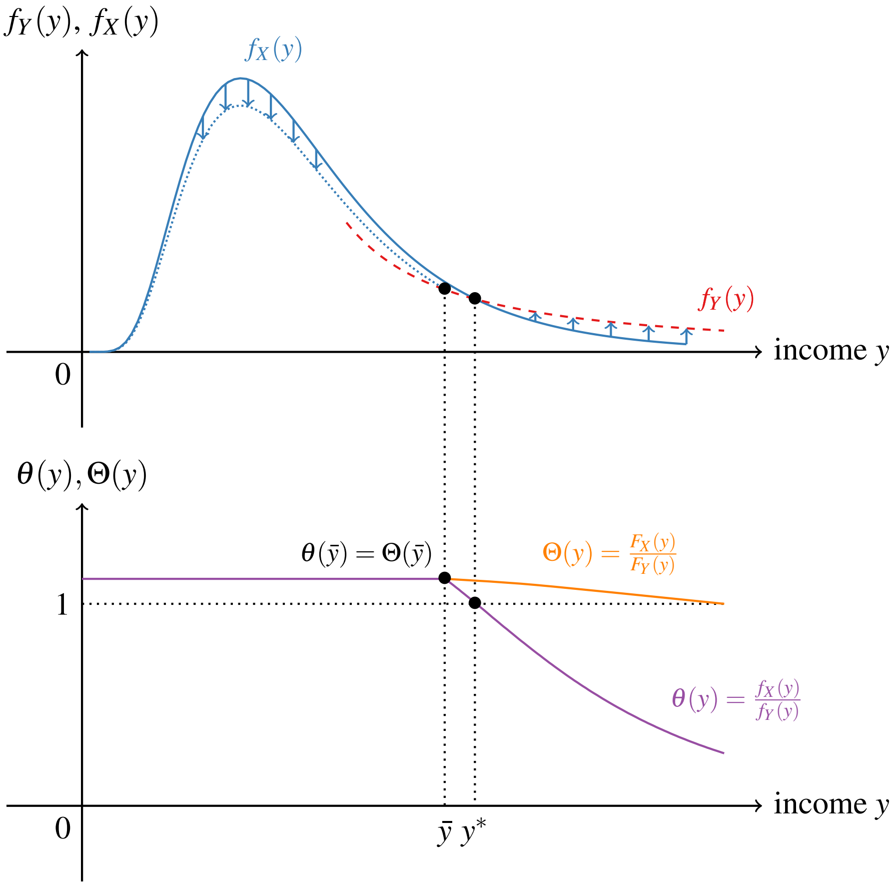


Figure 5.2: Choice of the Merging Point

The precise method for choosing the merging point seeks to preserve the continuity of the underlying income density function, and to realistically account for the form of the bias present in figure 5.1. In cases where the trustable span is large enough, such that the survey and tax distributions overlap, the procedure for endogenously choosing the merging point is presented in figure 5.2.

[4] The method is flexible enough to also confront cases where the trustable span may be too small to observe an overlap between the densities. For these, additional parameters need to be defined from neighboring years or countries to extrapolate the form of the bias and thus the correction to the whole distribution. See Appendix B in Blanchet, Flores, and Morgan (2019).

Let $\theta(y) = f_X(y)/f_Y(y)$ be the ratio of the survey density to the tax density at the income level y. This ratio is approximated empirically by comparing the frequencies in the percentiles of the interpolated tax distribution described in section 5.2.1 with the corresponding frequencies in the survey distribution. The value of $\theta(y)$ can be interpreted as a relative probability. If $\theta(y) < 1$, then people with income y are underrepresented in the survey. Conversely, if $\theta(y) > 1$, then they are overrepresented.

Note that what matters for the correction is the probability of response at a given income level *relative* to the average response rate. Intuitively, if some people are underrepresented in the survey, then mechanically others have to be overrepresented, since the sum of weights must ultimately sum to the population size. This basic constraint has important consequences for how we think about the adjustment of distributions. Any modification of one part of the distribution is bound to have repercussions on the rest. In particular, it makes little sense to assume that the survey is not representative of the rich, and at the same time that it is representative of the non-rich.

Let $\Theta(y)$ be the cumulative-density counterpart to $\theta(y)$, i.e., $\Theta(y) = F_X(y)/F_Y(y)$. Like $\theta(y)$, it can be estimated directly from the data. As figure 5.2 shows, in order to ensure a plausible form for $\theta(y)$, and for the density function to remain continuous, the merging point has to be chosen as the highest value of y such that $\Theta(y) = \theta(y)$. Tax data is discarded below this point. We call this point $\bar{y}$.

Since we must preserve the total population in the survey — which has to sum up to the total population — densities below the merging point must be lowered. Given the lack of tax data below the merging point, we choose to uniformly reweight observations below $\bar{y}$. This corresponds to the dotted blue line in figure 5.2. The implicit assumption made is that the relative probability of response stays constant over this part of the distribution, which seems to be satisfied in practice in countries with sufficient tax data (see Blanchet, Flores, and Morgan, 2019, section 3.2.2.).

The reweighting stage of the correction can be interpreted as addressing the survey's *nonsampling* error — that is, error from heterogeneous nonresponse rates across the population. Such a calibration procedure has several benefits. It allows users to adjust weights while respecting the aggregate consistency of other variables known to be already representative (e.g., age, gender). But it also allows users to incorporate other distributional dimensions from the tax data, such as population characteristics or income composition, if they exist, as well as dealing with differing income concepts. These additional dimensions are elaborated in the section 5.2.1.3.

After applying the calibration, the survey should be statistically indistinguishable from the tax data. This would be sufficient if the goal were to use the weighted survey to run regressions on the relationship between different variables. However, the precision that we get at the top of the income distribution may still be insufficient for some purposes, namely, accurately estimating inequality indicators, especially those that emphasize the top of the distribution, like top income shares. Indeed, they suffer from severe small-sample biases, which in general lead to an underestimation of inequality.

This *sampling* error — which comes from the limited size of the survey sample — can be overcome by expanding the survey's support. Practically, the method involves replacing the survey distribution beyond the merging point with the tax distribution, matching the survey covariates to it by preserving the rank of each observation. Statistically, this implies using the tax data for the marginal distribution of income, but keeping the survey data when it comes to other variables (and their dependency, by which we mean their copula, with income). This step ensures that the income distribution from the tax data is reproduced, while the survey's covariate distribution (including the household structure) is preserved. Although the socio-demographic totals for variables other than income are maintained, their conditional distribution can vary.

5.2.1.3 Extensions

The calibration procedure can improve the representativeness of the survey across several dimensions at the same time. The main point of the method is to ensure the survey's representativeness of top incomes using the tax data. But the procedure can also preserve or enforce representativeness of other variables, and can enforce representativeness in terms of the composition of taxable income, or the composition of the population by income bracket, if these statistics exist in the tax data.

First, there are variables for which the survey is assumed to be already representative. In this case, the distribution of these variables should be left unchanged by the calibration procedure. They typically correspond to gender and age groups, for which surveys have in general already been calibrated using census data.[5]

The second case corresponds to variables for which the survey is not currently representative, but for which we want to enforce representativeness based on external data. These variables and their distribution can be specified using a dummy variable in the code whose total average value

[5]The original distribution of these variables can be preserved using the option `holdmargins` in the `bfmcorr` Stata command. See the documentation of the command.

should be enforced by the calibration.

p	thr	women	employees
0.10	1000	0.60	0.90
0.15	2000	0.56	0.88
0.30	3000	0.53	0.89
...	...	...	...
0.90	10000	0.37	0.74
0.95	15000	0.31	0.71
0.99	20000	0.25	0.69

Table 5.2: Example of Tabulation with Demographic Characteristics

Use of Socio-demographic Characteristics by Income Sometimes, we have additional information on population characteristics by income brackets from tax data. Examples are the fraction of women or employees by bracket. These data should take the form presented in table 5.2.

Here, *p* refers to the percentile of total income in the tax data, *thr* refers to corresponding thresholds of the income variable, and the other columns refer the frequency of a given population characteristics in the tax data. For example, here, people earning between 1000 and 2000 represent 5% of the population. Among them, 60% of are women, and 90% are employees. Each column name other than *p* and *thr* should correspond to a dummy variable in the dataset equal to 1 if the observation belongs to the population category, and 0 otherwise.[6]

Use of Income Composition Data Another source of information that is commonly available in tax data is the composition of income within brackets. Using that information is useful if we assume that the bias may be different for people that derive their income from, say, capital rather than labor. This additional data should be inputted into the program in the form presented in table 5.3.

For example, here, 93% of the income of people earning between 1000 and 2000 in total income comes from labor, while 7% comes from capital income. Each column name other than *p* and *thr* should correspond to a variable in the dataset, otherwise it will be ignored with a warning. The intuition of this computation is that units will see their weight decrease or increase depending on whether their capital share is below or above the average of the bracket they belong to.[7]

[6] See Blanchet, Flores, and Morgan (2019) section 2.2.2., p.19, for formal details.
[7] Ibid.

p	thr	labor	capital
0.10	1000	0.93	0.07
0.15	2000	0.90	0.10
0.30	3000	0.88	0.12
...	...	...	...
0.90	10000	0.62	0.38
0.95	15000	0.53	0.47
0.99	20000	0.41	0.59

Table 5.3: Example of Tabulation with the Composition of Income

Tax Data with Partial Income Coverage Until now we have considered the case where the income recorded in tax data is a fairly comprehensive concept, which we can assume is likely to capture whatever truly drives the bias. Yet sometimes only a small part of income is recorded in the tax data. For example, in developing countries, only income from the formal sector may be recorded in the tax data, and there is a sizable informal sector only present in the survey data, which is widely spread across the distribution (e.g. Czajka, 2017).

In such cases, it would be problematic to directly apply the calibration method described previously. Indeed, since the adjustment factor for the weights would only depend on formal sector income, two people with the same income, one working in the formal sector and the other in the informal sector, would see their weight adjusted very differently. As a result, there would be almost no correction for the income distribution of the informal sector.

Therefore, if the income concept in the tax data is too limited, we can separate two different concepts: the (limited) one present in the tax data, and the (more comprehensive) one that is assumed to drive nonresponse. The formal way to deal with this problem is via the generalized calibration approach of Deville (2000). The method will construct a nonresponse function that depends on total income, while estimating it so as to match the income from the tax data.[8]

5.3 Distributing Tax-exempt and Other Unreported Incomes

Once we have constructed the microfile of income — either using tax data as the main source (as in section 5.1) or survey data (as in section 5.2) — the data can be assumed to be representative of incomes that were initially reported in the sources used for the microdata.

But in general, these datasets will still miss many types of incomes that are traditionally absent

[8]See Blanchet, Flores, and Morgan (2019), section 2.2.2, pp. 19–21, for formal details.

from tax and survey data. This can be because these income flows lie outside of the household sector (e.g., undistributed corporate profits), because they are imputed rather actual income flows (e.g., imputed rents, imputed social contributions), because they correspond to tax-exempt income flows (e.g., certain tax-exempt savings accounts), or because the microdata has some deficiencies (e.g., surveys that only report post-tax incomes).

The distribution of these incomes has to be imputed, either based on external sources, or using the same distribution as other types of income in the data. The precise way of doing so depends on the data available in a given context. Throughout the rest of this section, we try to distinguish "detailed" methods to apply with relatively high-quality data, from "simplified" imputations that can be used in more problematic contexts. The sections below provide general recommendations, but in each case the choice of the best approach is left to each researcher's judgment. Table 5.4 summarizes recommendations regarding pretax income flows for the both the simplified and the detailed case. Table 5.5 does the same for taxes and transfers.

5.3.1 Imputed Rents

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, this taxation has been abandoned in most countries. Therefore, we cannot observe imputed rental income in tax declarations, and we must use other sources. Note that even if administrative data does include imputed rents, tax authorities often value them according to estimates which are deliberate underestimations of actual market prices.

For survey data, the situation is mixed. Some of them do include imputed rents, but many do not. For example, the EU-SILC do include a variable for imputed rents in the latest waves, even though it is not yet included within the headline inequality statistics. But note that while the evaluation of imputed rents in surveys can be the result of rigorous estimates, in some cases they are much cruder — for example, simply defined as a fixed percentage of the household's income.

For simplified estimates, and whenever imputed rents are not reported in household surveys, they can be imputed to the distribution of consumption, or wages, or just follow an average by fractile based on their distribution in other countries. For detailed estimates, when surveys are the predominant microdata source, they can be imputed to households using hedonic rental price regression models estimated on rented dwellings with similar characteristic.

When tax microdata is the predominant source, different methods can be used (see also sec-

SNA concept (pretax)	SNA code	simplified	detailed
		imputed to	imputed to
Social insurance contributions	D61 (S14)	Wages	Simulations from information on gross incomes, rates and assessment ceilings
Imputed rent	Part of B2	Consumption / Wages / Average of other countries	Rental prices from hedonic rental price regression models / Mean-value imputation from surveys
Property income paid	D4 (S14, received)	Total income	Interest paid from microdata
Investment income disbursements	D44 (S14)	Capital income	
Insurance policyholders	*D441 (S14)*		*Wages*
Pension funds	*D442 (S14)*		*Wages*
Investment fund shareholders	*D443 (S14)*		*Corporate stock ownership/ Dividends + employer profit withdrawals*
Undistributed corporate profits	B5 (S11+S12)	Capital income	
Household share		*Capital income*	*Corporate stock ownership/ Dividends + employer profit withdrawals*
Government share		*Factor or pretax income*	*Factor income*
Net production taxes received by the government	D2 – D3 (S13)	Factor or pretax income (pretax)	Factor income (pretax)/ Consumption + imputed rents (post-tax)
Net property income received by the government	D4 (S13)	Factor or pretax income	Factor income
Social insurance surplus/deficit	D61 – D621 + D622 (S13)	Pretax income	Wages + pensions

Table 5.4: Imputing Missing Pretax National Income Components

tion 6.2). The distribution of imputed rents can be estimated using an imputation method relying on housing surveys (see Garbinti, Goupille-Lebret, and Piketty, 2018). Alternatively, we can rely on additional information reported in tax microdata such as property taxes to infer the value of imputed rents (see Piketty, Saez, and Zucman, 2018).

5.3.1.1 Property Income Paid

Property income paid by households comprises interest paid by homeowners, by landlords and by firms (D41) as well as land rents paid by households (D45). The integration of these items

SNA concept	**SNA codes**	**simplified**	**detailed**
		imputed to	**imputed to**
Production taxes	D2 – D3 (S13)	Factor or pretax income (pretax)	Factor income (pretax)/ Consumption + imputed rents (post-tax)
Corporate taxes	D5 (S11+S12)	Capital incomes	Corporate stock ownership/ Dividends + employer profit withdrawals
Personal income and wealth taxes	D5 (S14)		Simulations from legislated rates and thresholds/ tax liability from tax or survey data
Social assistance cash transfers	D623 (S14)		Simulation from eligibility rules/Means-tested transfers in surveys
Government final consumption expenditure	P3 (S13, paid)		Equal lump sum (health)/ Total income (other)

Table 5.5: Imputation of Taxes and Transfers

varies with the data source and quality. Interest paid by firms and landlords is often deductible from taxable income and thus is already implicitly integrated in tax micro data. In some surveys, one can find information on loan repayments including interest and repayment share. In this case, annual interest payment can be approximated using information on average interest rates and payment periods. If information on interest paid is missing, a proportional deduction from total income is the most suitable option.

5.3.1.2 Investment Income Disbursements

Investment income disbursements can be taken together and imputed to capital income in microdata in the simplified scenario. However, this strategy is likely to overly concentrate these income sources in fewer hands than is warranted by the nature of these incomes. A more precise approach — possible with the decomposition of this aggregate in the SNA — is to impute each component to an income type that best fits the constitution of the funds. Investment income attributable to insurance policyholders and pension funds should be imputed directly to formal wages, while income attributable to investment fund shareholders should be imputed to recorded dividends and employer profit withdrawals in surveys.

5.3.2 Undistributed Corporate Profits

In national accounts, undistributed profits correspond to the net primary income of corporations, net of corporate capital depreciation, where the primary income is the sum of retained earnings and the corporate income tax. The portion of these profits that should be redirected to households are those of corporations owned by resident households. The amount of (net) foreign retained earnings is already dealt with in the WID (see sections 2.1.1.1 and 4.1.2.3). This leaves us with the issue of the private and public share.

Various data sources could be used to estimate the government component of corporate undistributed profits, such as business surveys reporting the ownership of firms by institutional sector, or the balance sheets of the SNA. The latter can be used to determine the composition of corporate net equity liabilities (AF5) between households, government and foreign sectors. However, these sources are not always and everywhere available, particularly in less developed countries. In their absence, we suggest to use the government's share of property incomes received — that is, D4 (resources) of the general government, divided by D4 (resources) of the total economy — as a proxy for its share of undistributed profits.

Undistributed profits (or, to be precise, their personal component, see sections 2.2.1.1) should in theory be distributed in proportion to equity wealth. However, equity wealth is currently absent from virtually all income data sources. One solution is to import it from an external source that contains this information, such as available wealth surveys (e.g., Blanchet, Chancel, and Gethin, 2022). The other solution is to impute it to the most relevant observed income flow — in this case, dividends and profit withdrawals by business owners. When it is possible to match the individual profit of firms to the tax record of the owner (e.g., Alstadsæter, Jacob, et al., 2017), we should take advantage of this possibility, but this is very rare. In the simplified case, the overall distribution of capital income (or wealth) can be a reasonably good proxy for the distribution of this component (also see section 7.1.4).

5.3.3 Tax-Exempt Property Income

There is a variety of financial placements that exempt from the income tax. They include general-purpose savings accounts that are targeted at low- and middle-income households (e.g., the *Livret A* in France). They also include tax-deferred pension accounts. The income flows associated to these assets are by definition absent from most tax data.

In surveys, even when they are supposed to be included, they can be poorly captured. In particular, many small savings accounts represent a negligible source of income for their owners, so they often forget to declare the interest they earn from them. (This pattern is clearly visible in wealth surveys, in which it is common to see households that own such savings accounts yet report zero capital income.) The distribution of these financial placements — and their associated income flows — can be imputed from wealth surveys (see section 6.2).

5.3.4 Social Contributions

The treatment of social contributions in tax and survey data varies a lot. Recall that, in line with the SNA, we must account for both employees' and employers' contributions.

We sometimes observe employees' social contributions (in particular pension contributions) on individual tax returns, in which case this information should be used in order to obtain an accurate estimate gross wage income. Employers' contributions, however, are usually absent.

For surveys, and even when they report both pretax and post-tax income, the level of information at our disposal is generally not sufficient. Following recommendations by the Canberra Group (2011), the EU-SILC have pooled together employees' social contributions with taxes on income and wealth, while reporting employers' social contributions separately. This categorization is problematic in our view because there is (at least in principle) a clear, economically meaningful distinction between taxes and contributions (i.e., whether the system is insurance-based or not). On the other hand, the distinction between employers' and employees' social contributions is essentially a legal one. As for many other surveys, employers' contributions are usually not being reported at all.

Therefore when direct information is missing, we recommend that all available legislative and statistical information on graduated rates of social contributions by wage income and self-employment income levels be used to simulate social insurance contributions, both on the employers' and the employees' side. In developing countries, estimates should account for the fact that only formal-sector employees pay social contributions, implying that social contributions are generally more progressive than in high-income countries (see M. Fisher-Post and Gethin, 2023).

Regarding imputed social contributions (see section 2.1.2.1.1), these are by definition not observed. So we recommend allocating them in proportion to the wage incomes of the corresponding employees. In the simplified case, and in the absence of information, we can also resort to a similar rule for actual social contributions.

5.3.5 Direct Taxes on Income and Wealth

For personal income and wealth taxes, two possibilities can be used: we can apply simulations of direct taxes from legislative information on thresholds and rates, or we may use the information on tax liability that is often directly observed in surveys and income tax data (either micro files or tabulated). Ideally, we may try to combine both approaches. Where surveys are the predominant source, average effective tax rates by percentile calculated from the tax data can be deducted from corrected survey incomes or added to them if they are net-of-tax incomes. For the corporate tax, it should be distributed like undistributed corporate profits (see section 5.3.2).

Regarding inheritance taxes (and other personal taxes for which information is typically not available), and in the absence of direct data, we can adopt simplified assumptions. In the United States, we 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 information is available regarding the proportion of decedents and successors subject to tax and the progressivity of the tax (Piketty, Saez, and Zucman, 2018). Where wealth distributional data is not available, inheritance tax receipts can be imputed to top income earners. Similarly, for wealth taxes, they can be deducted from capital income earners.

5.3.6 Indirect Taxes

"Indirect" taxes such as property taxes — which can easily be construed as direct taxes, see section 2.1.2.1.1 — should be treated like direct taxes. For the rest, their distribution follows relatively simple principles exposed in section 2.2.2.2. Most components are to be imputed in proportion to pretax factor income. (Pretax post-replacement income can be used as an alternative in the simplified case.)

The most data-demanding task is the move from pretax to post-tax income, with respect to sales and VATs. Here, our convention is to remove these taxes from pretax income in proportion to consumption (see section 2.2.2.2). For that, we either make an estimation of consumption by income group using household budget surveys (Blanchet, Chancel, and Gethin, 2022), synthetic saving rates (Piketty, Saez, and Zucman, 2018), or other external information (Bozio et al., 2020). Another adjustment that should be made in low- and middle-income countries is that a significant share of poorer households' purchases are made in informal stores, which typically do not pay indirect taxes. This implies that consumption taxes are less regressive in these countries (see Bachas, Gadenne, and Jensen, 2024; M. Fisher-Post and Gethin, 2023). Countries with sufficiently precise data should make refined estimations that take into account how the typical consumption basket changes alongside the income distribution, and how these consumption baskets are affected by preferential rates or exemptions (see for instance Gethin, 2024c on South Africa).

5.3.7 Benefits

Cash and quasi-cash transfers. Regarding the imputation of social and means-tested cash transfers (D623, S14), we 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 matched income tax/household survey micro files such as the ERFS in France. In the case of the United States, we use the information on transfer receipts by income percentile, age and gender extracted from the CPS. In developing countries, many household surveys also report information on the coverage of social assistance programs and amounts received by each household, which can be used to allocate social assistance expenditure.

In-kind transfers and collective consumption. Regarding the allocation of public services, three types of complementary sources can be used. First, comprehensive administrative data can be used in some cases, for healthcare in particular, covering exact transfers (e.g., co-payments) received at the individual level (Bruil et al., 2025; Germain, André, and Blanchet, 2021). Second, household surveys often provide less granular but still useful information on intensity of use of public healthcare (e.g., health expenditure, visits to public healthcare providers, etc.), public education (e.g., school attendance of children in the household), and sometimes other services (e.g., use of public transport, fuel expenditure, beneficiaries of public housing programs, etc.). Third, government budget reports, other administrative sources, or other data sources can provide useful information on the geographical distribution of public expenditure, which can be used to account for spatial inequalities in transfers received.

5.4 Rescaling to Macroeconomic Aggregates

After the microdata has been completed with auxiliary information, we should have a more complete representation of income at the household-level. However, the aggregate level of income may still be significantly lower than the aggregate household-sector income from the SNA (with the obvious exception of components that were already imputed based on SNA data, see section 5.3 above). This may be due to conceptual differences in the definition of income, as well as to the persistence of measurement error in the microdata. Thus, we wish to correct the microdata so that it matches the aggregates from the SNA.

The concepts in the two datasets must be as close to each other as possible to avoid conceptual errors. Once this equivalence is achieved, we need to know what incomes to "match and rescale," which requires that we correctly link the income components in the corrected microdata to the SNA. This is crucial, because the rescaling of microdata incomes to macro totals can significantly affect the distribution. The outcome will depend on the level of income aggregation chosen for the exercise, a choice that is often constrained by the level of precision of the underlying data. In general, the narrower the level of aggregation the greater the distribution changes. This is due to the fact that some income components exhibit larger discrepancies with macro aggregates than others. And these components may be more unequally distributed at the micro-level (e.g., dividends). On the contrary, directly rescaling total income proportionally will only alter income levels and not the distribution.

Although the ideal level of income aggregation depends on country-specific details, table 5.6 presents a stylized approach to provide a reference. There can be a wide range of practical cases, according to the varying quality of both the micro and the macro data (see sections 5.4.1 and 5.4.2, and country-specific papers for details).

survey concept (pretax)	**rescaled to**	**SNA codes**
Wages (gross)	Compensation of employees	D1
Capital incomes (net) (interests, dividends, rental income)	Property income (net) − Retained earnings on FDI − Investment income disbursements	D4 − D43 − D44
Imputed rent	Operating surplus	B2
Self-employed income	Mixed income	B3
Pensions and unemployment insurance	Social security benefits in cash	D621
Other social transfers	Social assistance cash benefits + Other social insurance benefits	D623 + D622
Other non-regular incomes	Other current transfers	D7

Notes: This table only provides a stylized description of the rescaling procedure for reference, it should not be treated as a benchmark. The SNA aggregates in this table correspond to the "resources" side of the household sector (S14) account.

Table 5.6: Rescaling Micro Incomes to SNA Aggregates

5.4.1 Labor Incomes

In income microfiles, we usually observe three different variables for labor income: 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 (before any specific deduction or exemption).

For the pretax factor income series, we exclude replacement income (i.e., we set individual replacement income flows to zero in income tax micro files), and scale up fiscal wage income and self-employment income flows in order to match the national accounts totals for factor wage income (i.e., compensation of employees, D1) and factor self-employment income (i.e., net mixed income, B3n) used in the definition of factor income.

In the pretax national income series, we scale up fiscal wage income, self-employment income and replacement income in order to match the national accounts totals for pretax wage income, self-employment income and replacement income as defined in section 2.2.3. 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 section 5.3.4). In case we do not have information on the breakdown of social contributions by labor status (i.e., wage earners vs. self-employed workers, a breakdown that is not always available in national accounts), then social contributions (D61, S14, paid) should be deducted proportionally. We then scale up fiscal replacement income in order to match pretax replacement income (i.e., social insurance income, D621 + D622, S14 and subcomponents as defined in section 2.2.3).

5.4.2 Capital Income

Scaling capital incomes is usually more problematic than labor incomes. Some components like dividends, withdrawals and interest are reported relatively straightforwardly in both national accounts and microdata. For others — like imputed rents — we have to make assumptions. In many cases, the categories used in national accounts are different from those present in the microdata.

Most capital incomes are highly concentrated in micro datasets, and (especially when it comes to survey data) the total amounts declared in them can be very small compared to those from the SNA. For instance, it is normal for total dividends in a survey to be ten or eight times smaller than the corresponding aggregate in national accounts (for total household income, survey aggregates are mostly between a half and three quarters of SNA aggregates). Since, often, only a small fraction of survey respondents report receiving any dividends, scaling them separated from other capital incomes would brutally increase inequality, which would partly be due to a biased measurement. Therefore, depending on the situation, it can make sense to combine dividends with other types of capital incomes before scaling them up. Fiscal data often provides a remarkably better coverage of capital incomes (at least for those that are not exempt from taxes), that is why it is a crucial component in the DINA procedure. However, most of the time, aggregates from administrative records still need to be scaled up to reach SNA levels.

In surveys and tax data, actual rents from real estate are generally reported under a single variable, whereas in the SNA they are scattered across many items. One example is the rent from non-dwelling buildings, which is inseparable from self-employment income in the SNA, since it is included in the definition of mixed income (B3). Moreover, rents from natural resources are included as part of property income (D4), and rents from dwelling buildings are part of the operating surplus of households (B2), which renders it indistinguishable from imputed rents (though this breakdown differs between countries, see section 2.1.2.1.1). The latter item is also difficult to match to microdata because the corresponding definition is often tailored for a different purpose or based on basic approximations. Therefore, in many cases, some imputations based on the microdata must be made to match definitions.

5.5 Estimation of the Female Labor Income Share

Cross-country comparisons of gender inequality in labor income are often hindered by inconsistencies in income definitions, the treatment of part-time work, and the inclusion or exclusion of certain sectors (Ponthieux and Meurs, 2015). The female labor income share aims to overcome these challenges by leveraging harmonized household survey data from multiple sources, including the European Union Statistics on Income and Living Conditions (EU-SILC), the Luxembourg Income Study (LIS), the ILO Harmonized Microdata Repository, and national surveys from ten countries—specifically China, India, and eight West African nations. By synthesizing these data, the Female Labor Income Share provides a comparable measure of gender disparities in labor income across diverse national contexts.

The female labor income share is computed based on country-year aggregates of the female (male) wage (self-employment) income. We build this database in several steps.

1. Data selection:

    (a) First, we give priority to EU-SILC data, if available. The main exception is for countries with long-standing, high-quality national surveys that are integrated in LIS. While EU-SILC was specifically created to provide harmonized data across European countries, LIS draws on national surveys that, in some cases, predate EU-SILC and offer higher quality. For instance, Germany's SOEP survey (used in LIS) is substantially richer and longer-running than the more recent EU-SILC, and the same applies to countries such as Italy, Denmark, France, Great Britain, and Norway. For Serbia, we rely on ILO data instead of EU-SILC.

(b) For India, China, and eight Western African countries we use national survey data. For China, we use the national survey, CHIP, which presents challenges in the harmonization of income over time and particularly of individualizing household farm income. For a more detailed methodology, see the corresponding technical note (Gabrielli, Neef, and Robilliard, 2024).

(c) If EU-SILC or national surveys are not available, we draw on LIS data.

(d) If neither EU-SILC nor LIS are available, we use ILO micro data.

2. We augment the benchmark data sources with secondary choices when they offer a broader year range. For example, we supplement EU-SILC data with LIS or ILO data for years prior to 2003, and supplement LIS data with ILO data where applicable.

3. We interpolate the FLIS and aggregate female (male) wage (self-employment) income linearly between original data points. Some data sources do not provide continuous coverage for every year, which leads to gaps in the FLIS time series. In such cases, we linearly interpolate over the time dimension between two original data points to fill the gaps.

4. To extend the time series to cover 1990–2023 and to include countries missing from the earnings database, we use a two-step imputation approach. First, we estimate the female labor income share as a simple linear function of the female shares in wage- and self-employment and world region indicators using the combined LIS-EU-SILC-ILO database. Second, combining the estimated coefficients with ILO's employment series, we predict the female labor income share for all countries and years for which ILO modelled estimates exist. More precisely, we estimate the following regression model:

$$\begin{aligned}\text{Female Labor Income Share}_{ct} = \alpha + \beta\,&\text{Female Share of Wage Employment}_{ct}\\ +\,\gamma\,&\text{Female Share of Self Employment}_{ct}\\ +\,\delta\,&\text{World Regions}_c + \varepsilon_{ct}\end{aligned}$$

where c indicates countries and t years. The variable Female Labor Income Share$_{ct}$ is the female labor income share for country c and year t, the variables Female Share of Wage

Employment$_{ct}$ and Female Share of Self Employment$_{ct}$ measure the shares of women among all wage-employed and self-employed workers, respectively. World Regions$_c$ corresponds to fixed effects for institutional and cultural differences of nine world regions.[9] Observations are weighted according to population size.

5. For jurisdictions without any labor market related information due to a lack of consistent data sources, we provide the regional average.

[9] We classify countries into nine world regions: Asia (excl. China) comprising 32 jurisdictions, China, the Former Eastern Bloc (24 jurisdictions), Latin America and the Caribbean (43 countries & jurisdictions), Middle East and Northern Africa (20 countries), Northern America (4 countries), Oceania (16 jurisdictions), Sub Saharan Africa (48 countries), and Western Europe (28 countries).

Key Points

- Preferably, to compute income DINA, we construct microfiles that combine information from both tax and survey data.
- In countries with sufficiently exhaustive tax microdata, we only start from these files, and use survey data to add information about non-filers and certain tax-exempt incomes.
- In countries where the tax data is too limited (because they only provide tax tabulations or because the informal sector is too large), we use the tax data to correct the surveys at the top using the approach of Blanchet, Flores, and Morgan (2019).
- Then, we distribute remaining missing income components using the principles laid out in chapter 2.
- Finally, we rescale income components to SNA aggregates. Several levels of aggregation may be used depending on the quality and the precision of the data.

Chapter 6

Wealth Distribution Series

Measuring the distribution of capital income and wealth involves a large number of imperfect and sometimes contradictory data sources. In our view, it is critical to combine various sources of wealth data and methods to obtain a consistent picture of inequality. To estimate capital income and wealth, our preferred approach is to rely on the MICS method, which combines income tax data with household surveys and national accounts.

In this approach we start from income tax data and use the income capitalization method (section 6.1) to compute assets that generate taxable income flows. Because a number of important asset categories usually do not generate taxable capital income flows (section 6.2), it is always necessary to supplement the income capitalization method with imputations using household surveys, thereby making it a "mixed method" — using the terminology initially introduced by Atkinson and Harrison (1978).

The key contribution of this method is to allow researchers to overcome the drawbacks of using different data sources and methods separately. In section 6.3, we discuss how other data sources on wealth (including inheritance data and wealth rankings) should be reconciled and possibly combined with the income capitalization method. More details are provided in the country-specific papers (see in particular Bach, Bartels, and Neef, 2020; Garbinti, Goupille-Lebret, and Piketty, 2020; Martínez-Toledano, 2020; Saez and Zucman, 2016).

6.1 The Capitalization Method

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 (Atkinson and Harrison, 1978; Saez and Zucman, 2016) calculated using flows and stocks from national accounts. In more sophisticated versions, one can introduce different rates of return within each asset class, due to idiosyncratic variations in rates of return, and/or because the rate of return $r(k)$ tends to rise with the level of asset holding k (see the discussion below).

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 3.3).

Start with business assets and self-employment income. We simply assume that the ratio of self-employment income over 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. In the case of the United States, 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, but it is difficult to break down business assets into corresponding categories.[1] Hence, at this stage we apply the income capitalization method with a single category for self-employment income and business assets (Garbinti, Goupille-Lebret, and Piketty, 2020).

For housing assets, we usually observe effective rents in income microfiles (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 wealth or housing surveys or be directly observed in satellite housing accounts), we can scale up actual rental income in proportionally. From there we can estimate the value of tenant-occupied housing by dividing actual rental income by the average of return on housing (as computed in table 3.3). Imputed rental income (i.e., the rental value of owner-occupied housing) used to be

[1] "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.

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 6.2 below, where we also address the issue of household debt imputation).[2]

We usually observe a number of different categories of financial assets income in income tax microfiles. There are variations across countries, but generally we observe at least two categories — interests and dividends — which can be scaled up to the corresponding national accounts aggregates in a proportional manner. From there we can estimate interest-bearing assets (i.e., currency, deposits and debt assets) and dividend-bearing assets (equity and investment fund shares), using the categories defined on table 3.4.[3] The information that is available in income micro-files about income attributed to life insurance and pension funds is usually insufficient, so other sources must be used (see section 6.2 below). Generally speaking, the information available about financial asset income varies a lot across countries, and we recommend performing several sensitivity checks regarding the classifications about assets and rates of return that are being used to apply the income capitalization method.

6.2 Imputations of Other Assets and Liabilities

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 of asset holdings) consists of using either household wealth surveys and/or additional information from income tax micro files. 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 data sources. If the wealth survey contains sufficiently many observations, it is better to use a very flexible imputation method, i.e., one can estimate the percentage of homeowners and average home values for each cell defined by age, gender, percentile of labor income and percentiles of non-

[2] In Spain, imputed rents from primary residence has been exempted from personal income taxes since 1999, but imputed rents from other owner-occupied residences are still subject to the personal income tax (Martínez-Toledano, 2020). Hence, other owner-occupied housing is directly capitalized and not imputed.

[3] As the returns associated to debt assets and deposits may be very different, capitalizing them together could be problematic (Bricker, Henriques, and Hansen, 2018; Kopczuk, 2015). Whenever possible, it seems preferable to estimate separately debt assets from deposits (see the discussion in Garbinti, Goupille-Lebret, and Piketty, 2020).

housing wealth, and so on for other assets.[4] 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 Garbinti, Goupille-Lebret, and Piketty, 2020, for a detailed discussion on imputations based on household surveys).

Alternatively, one can rely on additional information reported in the income tax data that are highly correlated to the value of the asset to be imputed. For example, the value of owner-occupied housing can be inferred from property taxes and mortgage interest payments reported in the income tax micro files (e.g., Saez and Zucman, 2016).

6.3 Reconciliation with the Estate Multiplier Method

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 may 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 transparently.

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 (Piketty, 2011, see). In the United States, 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 shown that the estate multiplier method underestimates the rise of wealth inequality, as compared to both the income capitalization method and the SCF. In contrast, Alvaredo, Atkinson, and Morelli (2018) argue that British income tax microfiles (and also British national accounts and wealth surveys) make it difficult

[4] One drawback of this simple approach is that for a given cell, each asset holder holds exactly the same imputed amount. This limitation can be overcome easily by computing the different percentiles of home values for each cell.

to apply the income capitalization in a satisfactory manner — while inheritance tax data is more comprehensive than in the United States. Thus, they favor the estate multiplier method.

In the case of France, Garbinti, Goupille-Lebret, and Piketty (2020) find 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 microfiles, which are available since 1970 and offer exhaustive coverage of all income declarations in recent years, like in the United States). Again, 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.

In theory, the ideal data source to study the distribution of wealth would be high-quality annual administrative data on wealth, based on 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 excludes many assets).

Unfortunately, there are major limitations with the income 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 anymore and is not reported in income declarations. In particular, interest and dividend income tends to be taxed separately (sometimes 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. Moreover, even in countries where a substantial part of capital income flows are observable in income tax 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 6.2 above). Finally, the basic income capitalization method assumes a constant rate of return within each asset class, which may or may not be correct.

Our objective in the WID 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, mainly 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 — e.g., Atkinson and Harrison (1978), who apply both methods to the United Kingdom's income and inheritance tax data ranging from the 1910s–1920s up to the 1970s). However, in recent decades the two methods sometimes appear to deliver inconsistent results. Using income and inheritance tax data in the United States, 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 a very limited or even nonexistent 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 economies of scale 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 the WID 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 lose 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.

6.4 Wealth in Surveys and Tax Tabulations

Where tax microdata is unavailable, an alternative strategy to estimate the net wealth distribution is to combine wealth information from household surveys with that from summary tax data. The latter could take the form of tabulation information from the wealth tax or from wealth reported on income tax returns. As mentioned above, wealth taxes are not common across countries, and where they exist they tend to cover a small fraction of the population. In some countries, particularly those in Latin America, it is a legal requirement to report assets and liabilities when filing an income tax declaration, both in countries that have some kind of a wealth tax (such as Argentina and Colombia) and in countries that do not (such as Brazil). Where such information can be obtained, an approach similar to the one laid out for income in chapter 5 can be followed.

6.5 Global Wealth Distribution

For countries and years not covered by the previous methods, we can estimate the wealth distribution using a series of systematic, transparent imputations. Our basic method for the distribution of wealth rests on the observation that wealth inequality is highly correlated with income inequality.

We estimate the distribution of wealth in a given country i using a weighted average of the wealth distribution of other countries for which it is observed. To exploit the correlation of Figure 6.1, we give more weight to countries that have a similar level of income inequality.

We first normalize the wealth distribution to one in every country. Then, we calculate a_{itp}, the average normalized wealth of the g-percentile p in country i in year t as:

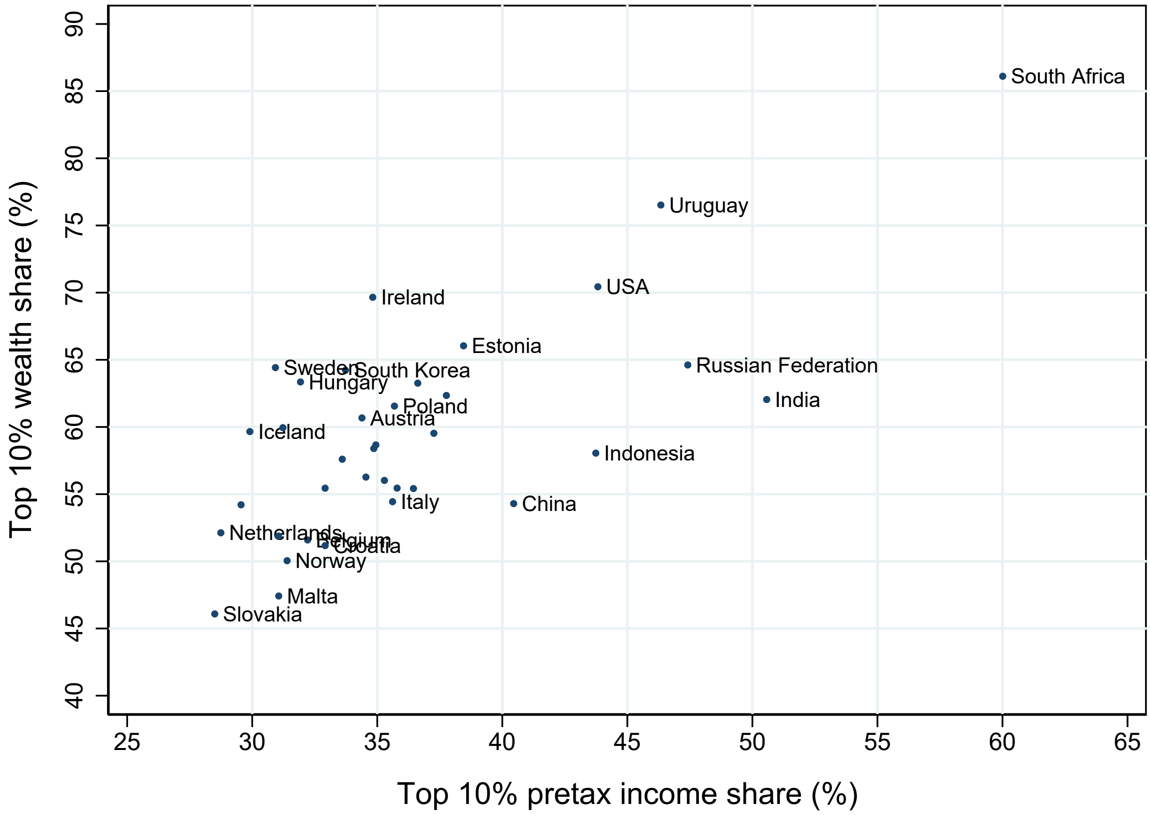
$$a_{itp} = \frac{\sum_{j,s} \frac{1}{h} K\left(\frac{r_{ti} - r_{jt}}{h}\right) a_{jsp}}{\sum_{j,s} \frac{1}{h} K\left(\frac{r_{it} - r_{js}}{h}\right)}$$

where $r_{js} \in [0,1]$ is the position of country j in year s in the ranking of all countries and all years in terms of top 10% income share, and where K is a standard Gaussian kernel.

The key parameter of that equation is the bandwidth h, which determines the degree of smoothing. We can use a leave-one-country-out cross-validation procedure to determine an optimal bandwidth.

The set of estimates produced at this stage—using either actual data or imputations—can potentially underestimate in some years the number of billionaires as compared to Forbes data. In order to recover the number of Forbes billionaires, we rescale the top of the distribution with Forbes data, assuming that aggregate wealth is unchanged and the distribution within the non-billionaire group is unchanged. By doing so, it is possible to produce series that are consistent with rich lists.

We compute the total net worth of billionaires (before Forbes correction) and the fraction of



Source: WID estimates

Figure 6.1: The correlation of income and wealth inequality

individuals above 1 billion dollars using cumulative distribution functions and tools available in the gpinter Pareto interpolation package[5]. We use the most recent exchange rate and price index available to express all country-series in market exchange rate dollars (or MER euros).

In cases where billionaire wealth we estimate is lower than Forbes, we add to the 127th g-percentile of each country the difference between the theoretical and Forbes total wealth of billionaires. We leave the wealth distribution series unchanged otherwise. In cases where we add the difference between Forbes and our actual series, the same amount is subtracted proportionately to each group of the bottom 99.999% to keep aggregate wealth constant and to keep the relative shares inside the rest of the distribution unchanged. In practice, the gap is usually small as compared to the total wealth of the bottom 99.999% and only has very minor impacts on the wealth of these groups.
The exact formula of the correction applied to each bracket average is the following:

$$a_{corr} = a \times \frac{1 - max(\frac{w_{Forbes} - w_{WID}}{n_{Forbes}}, 0) \times n_{Forbes}}{w_{tot}}, \tag{6.1}$$

Where a is the bracket average of each percentile, w_{Forbes} and w_{WID} are respectively the Forbes and our estimates of billionaire wealth, n_{Forbes} is the Forbes number of billionaires, and w_{tot} is aggregate wealth in the country. This formula takes out the added Forbes worth from the whole distribution. We then add the amount of wealth that was subtracted from the distribution back to the 127th g-percentile only using the following formula:

$$a_{corr} = a_{corr} + max(\frac{w_{Forbes} - w_{WID}}{n_{Forbes}}, 0) \times \frac{n_{forbes}}{pop_{tot}} \quad \text{if} \quad p = 99.999 \tag{6.2}$$

From this modified distribution, one can compute new shares and bracket averages after fitting them through gpinter again in order to ensure that the general fit of the distribution corresponds to the top correction. The corrected number of billionaires is computed using the same technique as before, multiplying the fraction of individuals beyond the billionaires threshold by country population in each year. We then aggregate the country series by re-ranking all g-percentiles in a single distribution, which allows us to obtain and document global trends.

This first correction yields better approximations of the global worth and number of billionaires,

[5] The distributions are fitted using generalized Pareto interpolation, for which an online tool (https://wid.world/gpinter/) as well as an eponymous R package have been designed. For details on the procedure, see Blanchet, Fournier, and Piketty, 2022.

but national and regional figures remain uneven (See columns 3-4 of Table 1). We thus proceed to a second correction, which takes place mostly within the 127th g-percentile instead of adjusting the whole of the distribution of each country. For each country, we compute various thresholds at 1, 10, 100 million as well as 1, 10 and 100 billion US Dollars. After computing the top average of wealth at these thresholds, we rescale the total wealth and number of billionaires to match Forbes data.

The second correction shifts wealth from lower brackets to the upper brackets in the case where billionaires were underestimated, while wealth is redistributed to the millionaires if the converse is true. The former case is of little effect on the general distribution, since wealth can be concentrated at its upper end instead without needing any change below (any increase in the top average of a group leaves untouched the lower brackets). The latter is of more consequence, because multimillionaires can usually spill outside of the last g-percentile, and an upwards reevaluation of their wealth induces a re-ranking problem in the top 1%. Fortunately, this issue is limited to a very small number of countries and of little magnitude, and can be solved cases by explicitly restricting changes to thresholds included in each country's 127th g-percentile. Overall, this allows for estimates of the very top of the distribution to be much closer to reality than initial series.[6]

[6] For a more detailed explanation of the methodology and the caveats associated with it, please see Bajard, Bauluz, et al., 2025.

Key Points

- Unlike the distribution of income, the wealth distribution is rarely observed directly in available administrative data. As for wealth surveys, their coverage remains more limited than for income, and the issues regarding the top of the distribution are even more critical.
- For these reasons, we resort to indirect methods that combine various sources in order to measure wealth. Our preferred method is the MICS method, which combines capitalized income flows from tax data with survey-based estimates for assets that do not generate taxable income.
- We stress the need to assess the sensitivity of estimates to various assumptions and to check how they can be reconciled with other sources and methods, such as the estate multiplier method.
- In practice, the best way of measuring the wealth distribution will depend on the type of data available in each country.

Chapter 7

Countries with Limited Data

The objective of the WID is to release homogeneous 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. The previous sections detailed the construction of such series for countries for which relatively good data is available, meaning in general access to survey microdata and either tax microdata or survey microdata. There are, however, a number of countries for which information is much more limited, and therefore additional adjustments and imputations must be made.

Many countries still fail to release tax data (even in tabulated form) which makes it extremely delicate to track the evolution of top incomes. And while household survey collection has improved significantly in most countries over the past decades, the situation remains contrasted. For instance, most countries in Africa carry out household surveys on a regular basis but the available data often comes with important limitations. The surveys are not collected annually (the average gap between two surveys is about 4 years). Even within a given country, the surveys often use different questionnaire designs which makes comparison difficult, not only across space — like for high and middle income countries — but also across time. The surveys collect information on consumption rather than income.[1] Finally, despite some progress, access to the microdata remains difficult.

This section presents various methods that can be employed to overcome data limitations. The

[1]Household surveys are also different across EU countries. Some data centers like the LIS (https://www.lisdatacenter.org/) try to collect and harmonize socio-economic micro datasets from upper- and middle-income countries and makes them available into harmonized income micro-database. As of today, very few low-income countries are covered by the project.

estimates that result from these methods are inherently more fragile than those that are based on direct information. We stress that the statistical methods that we describe here should never be viewed as substitute for hard data — or worse, as an excuse for not seeking better data. At the same time, we believe that making these adjustments and imputations are better than not doing them. Indeed, in this domain, doing nothing is doing something. It is fairly common, for example, to see direct comparisons between countries that measure inequality using consumption and countries that measure it using income. The implicit assumption — that consumption and income inequality are the same — of such a comparison is definitely not innocuous. Indeed, we know that consumption tends to be systematically higher than income at the bottom and lower at the top, and thus consumption inequality tends to be lower than income inequality. This is why we prefer to apply a correction, even if this correction is not fully satisfying in our view.

While making these corrections, and especially given the limitations of the original data, we stress the importance of using simple, transparent methods. Each correction step must address a clearly defined issue. The assumptions and parameters used in the correction must explicitly correspond to well-identified phenomenons. In practice, given the lack of data, these parameters often have to be inferred from what we observe in other countries. This is never ideal, and those assumptions will be revised as better data become available. But this less-than-ideal situation is precisely why we feel that it is important for these assumptions to be made openly and transparently.

7.1 Countries with Tabulated Survey Data

7.1.1 From Survey Tabulations to Survey Distributions

For countries where access to microdata is difficult, the primary source of data consists in survey tabulations. Many of them are available from the World Bank, which makes them publicly available on PovcalNet.[2] Additional tabulations are included in the WIID, maintained by the UNU-WIDER. These tabulations provide information on the distribution of income or consumption per capita for various brackets. We recommend using the generalized Pareto interpolation method developed by Blanchet, Fournier, and Piketty (2017) (see section 5.2.1) to create a full distribution by g-percentile from these tabulations.[3]

The same can be true for historical time periods in both developed and developing countries:

[2] See `http://iresearch.worldbank.org/PovcalNet/povOnDemand.aspx`.

[3] Note that in many cases PovcalNet tabulations are already interpolated using parametric methods.

we sometimes use tabulated survey data, with cautious adjustments such as those listed below, to retrieve the full income distribution, provided that the survey tabulations meet the basic requirements of the Blanchet, Fournier, and Piketty (2017) method.

7.1.2 From Survey Consumption to Survey Income

One issue when using survey tabulations in developing countries is that they rely almost exclusively on consumption. This makes systematic comparisons between developed and developing countries difficult, since inequalities of economic resources in developed countries are more often measured by pretax or post-tax income. Income inequalities tend to be higher than consumption inequalities, since top earners tend to have higher saving rates, and income has a transient component that some households are able to smooth in order to maintain a stable level of consumption.

In order to make consumption and income inequality estimates comparable, we suggest a simple adjustment method using a logistic function. The idea is to exploit the relationship between consumption and income for each g-percentile, as observed in countries for which both distributions are available. This method (see Chancel, Cogneau, et al., 2019) is presented below.

Note that the objective is to make consumption and income inequality estimates comparable, not to model the relationship between income and consumption at the individual level. For that purpose, it suffices to know to what extent consumption is higher or lower than income for each percentile. Then, we can exploit this relationship to transform consumption inequality estimates into income inequality estimates. In other words, our aim is to model income-consumption profiles $c_1(.)$ of the form:

$$c_1(p) = \frac{Q^I(p)}{Q^C(p)}$$

Where $Q^I(.)$ is the quantile function associated with a given distribution of income, $Q^C(.)$ is the quantile function associated with a given distribution of consumption, and $p \in [0,1]$. If, given $c_1(p)$, we only have access to a consumption distribution $\hat{Q}^C(p)$ (we can then impute a corresponding income distribution defined by $\hat{Q}^I = c_1(p)\hat{Q}^C(p)$.

We start by looking at the empirical shape of $c_1(p)$ for all percentiles in countries and years for which we have reliable data on both income and consumption. Following our definition of $c_1(p)$, computing income-consumption ratios is straightforward: it simply consists in dividing the bracket average of each percentile of the income distribution by its consumption counterpart. In order to make profiles comparable, we systematically normalize average income or consumption to 1. Notice that since our aim is to use $c_1(p)$ as a multiplicative factor, the ratio of aggregate consumption to aggregate income is irrelevant: what matters is how $c_1(p)$ varies with p.

In countries for which we have information (Chancel, Cogneau, et al., 2019), we find that the relationship between income inequality and consumption inequality is distinctively S-shaped: average income is generally substantially lower than average consumption for the bottom quintile of the population. The ratio of income to consumption then increases more or less linearly up to percentiles 80 and 90, before rising dramatically at the top of the distribution. This is consistent with the mechanisms outlined above: poorer individuals tend to smooth their consumption, while the very rich tend to save a significant proportion of their current earnings. As a result, consumption inequalities are generally lower than income inequalities.

To reproduce this profile, we estimate $c_1(.)$ parametrically by using a scaled logit function of the form $(0 < p < 1)$:

$$c_1(p) = \alpha + \beta \log\left(\frac{p}{1-p}\right)$$

where α is a constant which determines the overall level of the curve. It is completely irrelevant to our imputation problem, since multiplying the quantile function by α only affects the overall mean of the distribution. Our parameter of interest is β: it determines how fast the ratio of income to consumption increases with p and is therefore a direct measure of the extent to which consumption inequalities are lower than income inequalities.

Using the estimates obtained for countries for which we have information on both income and consumption, we generate theoretical profiles for countries for which consumption alone is available. We use that profile to generate income distributions.

7.1.3 From Survey Income to Fiscal Income

The correction above lets us have income distributions that are comparable to those observed directly in surveys. But does not address the fact that income itself in underestimated at the top in surveys. Therefore, we apply a second correction that consists in correcting the average income of top earners.

Following the method used for consumption, our aim is to use existing data to define "plausible" profiles correcting income levels at the top of the distribution. We look at variations in the underestimation of top incomes in Africa by bringing together surveys and fiscal tabulations from Côte d'Ivoire (Czajka, 2017) and South Africa (Alvaredo and Atkinson, 2010). We then apply the method of Blanchet, Flores, and Morgan (2019) to combine surveys and tax data in order to get corrected survey income distributions (see section 5.2).

Exactly as in the case of consumption and income, our objective is then to study "survey-fiscal" profiles $c_2(.)$ of the form:

$$c_2(p) = \frac{Q^F(p)}{Q^I(p)}$$

where $Q^I(.)$ is the quantile function associated with the distribution of income observed in the survey, and $Q^F(.)$ is the quantile function of the distribution obtained after correcting for the under-representation of top incomes. Once again, both distributions are normalized so that their mean is equal to one.

We then estimate these profiles for countries for which we have information. The correction profile of top incomes can be formally conceptualized as depending on two parameters: the size of the group which is corrected, and the strength of the correction. One way to formulate these two dimensions parametrically is to model survey-fiscal profiles by the quantile function of the Lomax (or Pareto type II) distribution ($0 < p < 1$):

$$c_2(p) = \mu + \sigma(p^{1/\gamma} - 1)$$

where μ is a constant which determines the overall level of the curve; as in the case of consumption-income profiles, it is irrelevant to our problem. Since it makes sense to let $c_2(p)$ take 1 before a certain percentile p_0, one can set $\mu = 1 + \sigma$, so that $c_2(0) = 1$ and:

$$c_2(p) = 1 + \sigma p^{1/\gamma}$$

where σ is the scale parameter. It controls the slope of the curve: the higher σ, the higher top incomes are underestimated by surveys. The shape parameter is γ: as it decreases, the slope becomes more convex, so that a smaller fraction of top incomes is corrected. We then apply these "survey-fiscal" correction profiles to the income surveys to approximate the level of inequality we would expect to see in tax data.

7.1.4 From Fiscal Income to National Income

Under the assumption that our method for improving the measurement of income inequality is correct, the resulting dataset we obtain corresponds more or less to the distribution of household income — that is, the sum of compensation of employees, mixed income and property income received by the household sector in the national accounts.

To reach the national income and obtain figures on individual incomes which are consistent with macroeconomic growth, we have to make assumptions on the distribution of unreported income components. These mainly include the taxes on production received by the general governments and the retained earnings of corporations, which can constitute a significant fraction of the national income in both developed and developing economies. Because taxes on production are distributed proportionally to pretax income in the DINA framework, they do not have any impact on inequality. So the most important components are tax-exempt capital incomes such as the retained earnings of corporations.

In practice, even capital income components from the household sector are underrepresented in both survey and tax data, so that we could decide to add a fraction of other capital incomes (i.e., other than retained earnings) to the distribution. In practice we refrain from doing so, unless we can point to clear evidence that capital incomes are missing (for example if they are excluded from the income tax). But it means that our estimates of missing income are still probably conservative.

Depending on the type of data available, we can use different methods to account for missing incomes. Of the methods presented below, the first one (proportional allocation) does not require any data or imputation. The next two ones assume that we have an estimate of the marginal distribution of missing incomes (e.g., based on wealth, see section 7.3), but that we need to impute the dependency between missing and nonmissing incomes. The last one suggests directly imputing the amount of missing income that accrues to various groups.

In practice, our recommendation is to use the method described in section 7.1.4.2. In many cases, some of the parameters required to apply the method will be missing: either the marginal distribution of capital incomes, the dependency between capital and labor, or the total amount of retained earnings to distribute. In all of these cases, we provide default parameters based on countries for which we have the data. They are provided in the online appendix of these guidelines. These are to be used as a last resort: whenever possible, we recommend that researcher look for available data, or make their own estimations.

Depending on how developed the national accounts of various countries are, we may or may not have direct estimates of retained earnings. We encourage researchers to look for the data when it exists or, when possible, make their own estimates. But as a last resort, we offer default parameters.

7.1.4.1 Proportional Allocation

The simplest way of distributing missing income components is to distribute them proportionally, i.e., to directly rescale the distribution of fiscal income (or corrected survey income) to national income.

This approach was used in Africa (Chancel, Cogneau, et al., 2019) where national accounts are still in their infancy, and therefore do not provide reliable data on unreported income. This approach has also been used to produce estimates of pretax national income inequality based on raw fiscal income shares from the WTID for a handful of countries where more detailed work has not been undertaken yet.

This approach keeps the overall distribution unchanged, while making average incomes and growth rates more comparable across countries and over time. We stress that this is a conservative assumption: in most existing distributional national accounts studies, the imputation of unreported income leads to higher inequality levels, mainly because retained earnings are concentrated at the top the distribution (Blanchet, Chancel, and Gethin, 2022; Garbinti, Goupille-Lebret, and Piketty, 2018; Jenmana, 2018; Morgan, 2017). As better national accounts data, survey microdata and tax data become more systematically available, such estimates should be improved.

7.1.4.2 Reduced-form Estimation of the Distribution of Missing Incomes

To provide refined estimates of the distribution of national income, we have to impute how much income that is missing for survey/fiscal estimates accrues to the different income groups. To simplify the exposition of the problem, assume that our survey/fiscal estimate captures only labor income, and that the missing incomes we wish to distribute are capital income (the idea extends easily to more subtle settings).

How much capital income accrues to the different income groups depends on two factors: the level of capital income inequality (i.e., the marginal distribution of capital income) and the dependency between capital and labor income. Let us assume that we have an external estimate of the marginal distribution of capital income (based for example on the distribution of wealth, see section 7.3). How can we infer how much capital income should be attributed to each group?

A simple and transparent solution is to consider the following function ($0 < p < 1$):

$$\lambda(p) = \operatorname{asinh}\left(\frac{Q(p) - Q_L(p)}{\bar{Q}_K}\right) - \operatorname{asinh}\left(\frac{Q_K(p)}{\bar{Q}_K}\right)$$

where $Q(p)$ is the quantile function of total income, $Q_K(p)$ is the quantile function of capital income, $Q_L(p)$ is the quantile function of labor income, and $\bar{Q}_K$ is average capital income. If labor and capital incomes are perfectly correlated, then $\lambda(p) = 0$. Otherwise, we will usually have $\lambda(p)$ decrease from a value above zero to a value below zero. The hyperbolic sine function is meant to have the function behave like an absolute difference for small values and like a ratio for high values, and therefore does not introduce singularities because of zero capital incomes. The normalization by $\bar{Q}_K$ ensures scale invariance.

Assume that we know the function $\lambda(p)$, the marginal distribution of capital income $Q_K(p)$, and the marginal distribution of labor income $Q_L(p)$. Then, for the distribution of total income, we get:

$$Q(p) = Q_L(p) + \bar{Q}_K \sinh\left[\lambda(p) + \operatorname{asinh}\left(\frac{Q_K(p)}{\bar{Q}_K}\right)\right]$$

The function $\lambda(p)$ may be inferred from countries where the data allows us to estimate it directly. As shown in figure 7.1. In France and in the United States, we find that this function is generally decreasing and comprised between -0.5 and $+0.5$. The profile differs slightly between the two countries, but is very stable over time. We apply the average of these functions to other countries in order to distribute missing capital incomes in countries without sufficient data. We provide the corresponding data series in the online appendix.

7.1.4.3 Copula-based Methods

Another solution to infer the share of missing income that accrues to each part of the distribution is to use copulas. Copulas are mathematical objects that correspond to the joint distribution of the rank between two distributions: they are very useful to characterize the dependence between two distributions independently of their marginal distribution. If we assume once again that we know the marginal distribution of missing incomes, and that we assume a given copula between labor and capital income, then we can estimate the distribution of total income. The `gpinter` R package and online interface (`wid.world/gpinter`) does allow users to perform such operations using a specific parametric copula known as the Gumbel copula. The DINA estimates for the Middle-East (Alvaredo, Assouad, and Piketty, 2019) used this approach with a Gumbel copula parameter of 3 to account for missing incomes.

The copula-based approaches are useful in certain contexts (in particular is some form of microdata is available), but also more difficult to use and less transparent than the simple approach described in section 7.1.4.2. In general, we would recommend the simpler method, unless circumstances justify the use of a copula-based method.

Source: WIL staff estimates using data from Garbinti, Goupille-Lebret, and Piketty (2018) for France and from Piketty, Saez, and Zucman (2018) for the United States. See online appendix for construction and data series. *Note:* The values reported correspond to averages by g-percentile over all available years.

Figure 7.1: $\lambda(p)$ for France (1970–2014) and the United States (1962–2014)

7.1.4.4 Share of Missing Income Received by Percentile Groups

The two sections above were about combining data on the dependency between labor and capital income with data on the marginal distribution of capital incomes. But in some cases, it can be more straightforward to make direct assumptions on the share of missing income received by each g-percentile based on what is observed in other, similar countries.

This method was for example applied by Piketty, Saez, and Zucman (2019) to produce simplified DINA estimates for the United States that are similar to the complete ones, but the make the key assumptions simpler to understand, and the key results more transparent. This approach is, in particular, easily applicable to countries that provide the decomposition of the income of different groups by income types is their tax data — a relatively common occurrence.

However, outside of such specific cases, we recommend using the approach of section 7.1.4.2.

That is, even if we have to impute both the dependency between labor and capital, and the marginal distribution of capital incomes, we believe that it is better and more transparent to make both of the parameters appear explicitly, rather than make an assumption about their combined value.

7.1.5 Adjusting the Statistical Unit

Survey tabulations use specific statistical units that often differ from the one DINA series use. For example, PovcalNet uses per capita income or consumption (i.e., including children) while we only consider adults. Other tabulations may use the variety of equivalence scales that have employed in the literature (e.g., the square root or OECD equivalence scale).

In such cases, it is usually preferable to adjust distributions using a similar method as the one developed above to correct for the difference between income and consumption or survey and fiscal income.[4] Note, however, that differences between equivalence scales are usually more limited than differences between income and consumption, or survey and fiscal income.

7.1.6 Inferences within the Top Tail

Another difficulty of estimates based on survey data is that inference about narrow top income groups is made difficult by the fact that the underlying survey data provides very few observations in that part of distribution. This is a sampling error issue, which is separate from the nonsampling errors (e.g., misreporting and nonresponse) that we usually invoke to justify the underrepresentation of top incomes. Note that not all survey-based tabulations are affected by this issue. The tabulations from PovcalNet, for example, are in many cases estimated via parametric methods, which already solves this issue.

But for other estimates, it may be preferable to make another adjustment to make sure that the estimates of income at the very top (e.g., top 0.1%) are reasonable and sufficiently stable. To that end, Blanchet, Chancel, and Gethin (2022) fit a generalized Pareto model within the top 10% of the income distributions that they estimate. That is, they assume the final income distribution within the top 10% follows the generalized Pareto distribution, which has the following cumulative distribution function:

$$F(x) = 1 - \left\{1 + \xi\left(\frac{x-\mu}{\sigma}\right)\right\}^{-1/\xi}$$

[4]Blanchet, Chancel, and Gethin (2022) use a more sophisticated approach in Europe using a machine learning algorithm. In most cases, however, and given the data that is available, more simple methods should be sufficient.

This is a fairly general model that includes the Pareto distribution, the exponential of the uniform distribution as special cases. It is widely used in extreme value theory, which shows that it works as a universal model of distribution tails. The suggested estimation method is that of probability-weighted moments (Hosking and Wallis, 1987). That is, define $a = \mathbb{E}[X]$ and $b = \mathbb{E}[X(1-F(x))]$. Then we have $\xi = (a-4b+\mu)/(a-2b)$ and $\sigma = (a-\mu)(2b-\mu)/(a-2b)$, while μ is determined *a priori* from the threshold from which we start to use the model (e.g., the top 10% threshold). The complete distribution combines the empirical distribution for the bottom 90% with the generalized Pareto model for the top 10%.

One of the key advantages of this approach is that while it changes the distribution within the top 10%, it does not change the top 10% share or average, which at this stage are considered as reliable as possible given the underlying data.[5]

7.2 Combining Survey Tabulation with Tax Data

The previous section was concerned with situations where no direct tax data is available. Yet there are situations with poor access to survey microdata, but in which we have some access to tax data, either in the form of tabulations or in the form of top income shares from previous studies. This section describes how handle such cases.

7.2.1 Adjusting Tax Data

One issue with tax data tabulations is that they use a statistical unit (either individuals or tax units) that is different from the ones we use in DINA (usually equal-split adults). When we have access to survey microdata (section 6), our recommendation is to match the statistical unit in the survey to that of the tax data when applying the survey correction method of Blanchet, Flores, and Morgan (2019). Because the method preserves the structure of the microdata, it remains possible to change the statistical unit after the adjustment. This is a powerful way of overcoming the statistical unit problem, but unfortunately it only applies to situations in which survey microdata is available.

When no direct microdata is available, we recommend making simple adjustments based on

[5]Therefore, this is fundamentally different from attempts at correcting the underrepresentation of the rich using Pareto-type models. We generally refrain these methods because, in our experience, they behave too erratically to be reliable, and it is easy to reach virtually any level of inequality by arbitrarily choosing different thresholds for modeling the tail. They also have rather poor theoretical justifications. What we do here is only about improving the distribution within the top 10%.

correction profiles observed in similar countries with better data, as was done above to move from consumption to income, and to survey income to fiscal income. That is, use survey microdata to determine the ratio:

$$c_3(p) = \frac{Q^{EQ}(p)}{Q^{TAX}(p)}$$

where $Q^{EQ}(p)$ is the quantile function of the equal-split distribution, and $Q^{TAX}(p)$ is the distribution of the tax unit distribution (individuals or couples in countries with joint taxation). Make sure that both distributions have their average normalized to one to make comparisons easier. Then, apply the same ratio to the tax data to get a fiscal income distribution that uses the desired statistical unit.

7.2.2 Adjusting Survey Tabulations

Assuming that income concepts and statistical units between the survey data and the tax data have been made comparable, survey tabulations can be calibrated on the tax data using the same method from Blanchet, Flores, and Morgan (2019) that was used for survey microdata.

Concretely, a complete survey tabulation can be treated as microdata: each bracket works as one observation, with a weight equal to the number of people of the bracket. For that abstraction to work well, it is better to have a large number of brackets (at least every g-percentile, and possibly more) with can be obtained by interpolating the distribution with the method of Blanchet, Fournier, and Piketty (2017) (see section 5.2.1). The method of Blanchet, Flores, and Morgan (2019) can then directly be applied directly to such data. The output can be interpreted as a new tabulation that can be interpolated again using the generalized Pareto interpolation method of Blanchet, Fournier, and Piketty (2017) to generate the desired g-percentiles (see section 5.2.1).

7.2.3 Combining Survey Tabulations with Top Shares

To combine survey tabulations with tax data as outlined above works well if we can be relatively confident in income levels (not just the shares) of the survey data. But in some cases, the situation is more complicated. In countries with poor data, the survey might have to be corrected for both the concept (income vs. consumption) and the statistical unit, making the final result too reliant on fragile imputations.

In such cases, it might be preferable to compute top income shares directly by combining tax data for top incomes with a denominator for total income based on national accounts (or to use such an estimate from a previous study, which were generally published in the WTID, the precursor of the WID). And, then, combine the survey distribution with this top share estimate. One advantage of this approach is that it can ensure consistency with previous top income share studies. It is also insensitive to scaling issues.

Here, it is important to note that the correction method of Blanchet, Flores, and Morgan (2019) cannot be used, because is not meant make a survey match specific top income shares. It is meant to make a survey match a given level of income at the top, from which the top share is endogenously deduced. Therefore, to combine the survey with top income shares directly, we have to use different tools.

Our recommendation for doing so involves "stitching" together the Lorenz curve from the survey at the bottom with the Lorenz curve from fiscal data at the top. Assume, to simplify exposition, that we have top income shares from tax data for the top 10% (i.e., above the percentile $p_0 = 0.9$) of the distribution, and that we have a survey that we trust for the bottom 90% of the distribution.[6]

Let s be the value of the top 10% income share. Let $L_S(p)$ and $L_T(p)$ be the value of the Lorenz curves for the survey and the tax data at percentile p (hence, $L_T(p)$ is only observed for $p > p_0$ and $L_T(p_0) = 1 - s$). The first step for putting the two distributions together is to rescale the survey Lorenz curve $L_S(p)$. Define the adjusted survey Lorenz curve as $L_S^*(p) = (1-s)L_S(p)/L_S(p_0)$. This adjusted Lorenz curve defines a distribution in which every percentile as the same share of income within the bottom 90% as it does in the survey. We then typically arrive at a situation such as the one described in figure 7.2. The red curve at the top-right corner is the tax Lorenz curve $L_T(p)$, and the blue curve on the bottom left is the rescaled Lorenz curve $L_S^*(p)$. Together, they form the Lorenz curve for the new, combined distribution.

By construction, the overall Lorenz curve is continuous, and increasing. Observe, however, that around the point of juncture A, the Lorenz curve has two problems. First, it is not differentiable. The tangent to the red curve at A is the line (AB), while the tangent to the blue curve at the same point is (AC). The non-differentiability at the juncture implies that the quantile function at this point is discontinuous.

But even more problematic is the fact that the curve is non-convex at this point.[7] A nonconvex Lorenz curve implies that the underlying quantile function is nonincreasing at this point. Indeed, given the tax data Lorenz curve in red, the Lorenz curve for the bottom 90% of the distribution

[6] Here we assume that income concept and statistical units between both sources have been harmonized.
[7] This may not always happen, but it can.

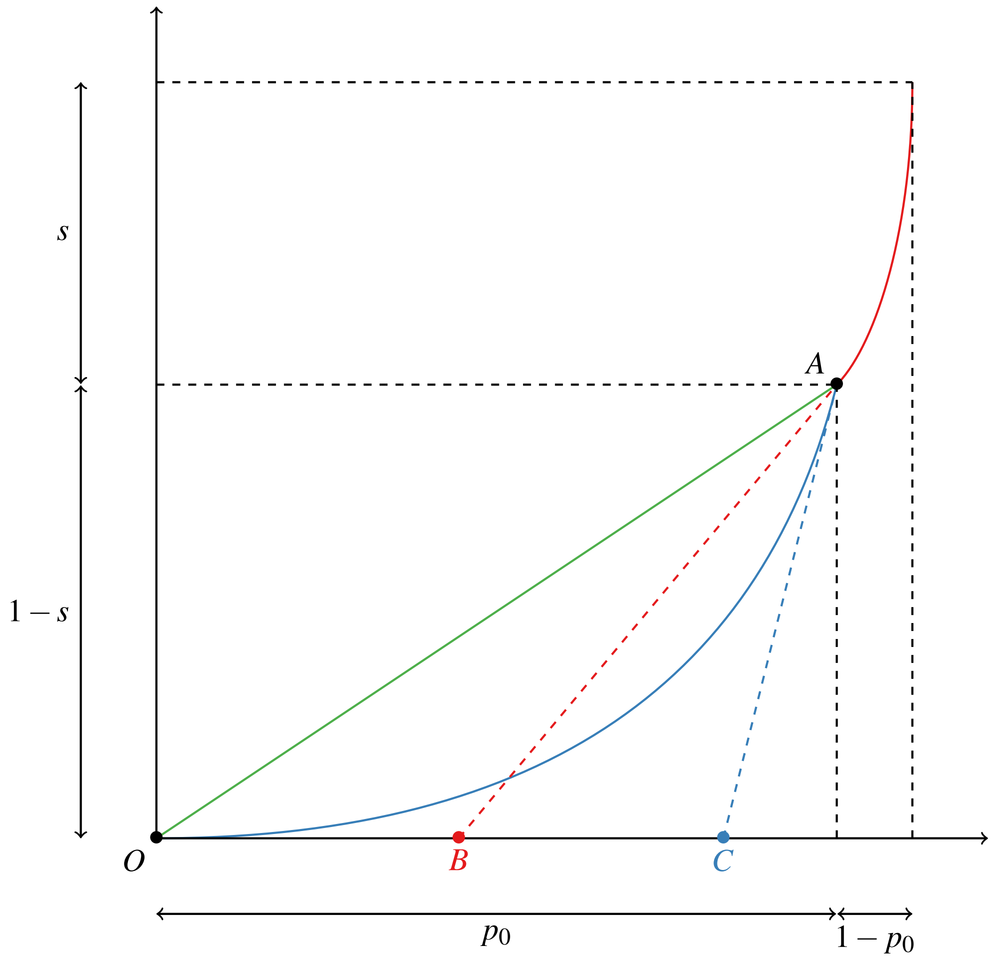


Figure 7.2: Stitched Lorenz Curves

must be comprised within the triangle AOB. The line (AO) is the line of perfect equality within the bottom 90% (everybody has the same income), while the line (AB) is the line of maximum inequality (the smallest group of people possible, i.e., the bracket $[B, p_0]$, has the same income as the one that defines the top 10% threshold). In figure 7.2, the blue curve exceeds the line of maximum inequality.

This issue is important to keep in mind because it speaks to a fundamental mathematical constraint: namely, it may well be impossible to define a globally consistent distribution that has the same shares as the survey data within the bottom 90%, and the same shares as the tax data for top 10%. Something has to give.

So, we have to make additional adjustments. We present two possibilities below, which make different trade-offs. The best one depends on what are the priorities in a given context. The online appendix includes a Stata implementation of both methods.

7.2.3.1 Method 1: Re-ranking

A first possibility is to re-rank percentiles the brackets of the stitched distribution (i.e., to sort them in increasing order). By definition, the resulting distribution will have an increasing quantile function. Moreover, as long as re-ranked brackets have identical sizes, the average of the distribution will remain the same.

This procedure is very simple to implement. It has the property of not modifying either the very top or the very bottom of the distribution. However, it will change the top 10% share (and, as a result, the bottom 90% share) as measured by the tax data. It also offers no real guarantee that the resulting quantile function will be continuous (and in fact, it generally will not be).

7.2.3.2 Method 2: Constraining

A second procedure involves constraining the Lorenz curve of the bottom 90% to be consistent with the tax data. This approach as the advantage of perfectly preserving the fiscal data at the top, and it ensures that the resulting quantile function is continuous. But as a result, it creates stronger distortions for the bottom 90% of the distribution.

Geometrically, the procedure works as follows (see Figure 7.2). For each value $L_S^*(p)$ of the survey Lorenz curve in red, we replace the corresponding percentile p by its value multiplied by the ratio of the horizontal distance (at the y-coordinate $L_S^*(p)$) between the lines (AO) and (AB) to the horizontal distance between the line (AO) and (AC). This has the effect of distorting the red curve to make it fit into the triangle AOB.

Algebraically, the computation works as follows. Start from the tabulation of the stitched, unconstrained Lorenz curve. That tabulation contains, for a series of percentiles p (i.e., the x-coordinate of the Lorenz curve), the bottom share $L(p)$ (i.e., the y-coordinate of the Lorenz curve), where:

$$L(p) = \begin{cases} L_S^*(p) & \text{if} \quad p < p_0 \\ L_T(p) & \text{if} \quad p > p_0 \end{cases}$$

To that tabulation, and for $p < p_0$, we will add three variables that associate to every y-coordinate $L(p)$ the x-coordinate of the lines (AO), (AB), and (AC). For the line (AO), simply define:

$$x_{AO}(p) = pL(p)/L(p_0)$$

For the lines (AB) and (AC), we need to know their respective slopes, i.e., the value of the normalized quantile at p_0 for the tax data and the survey, respectively. That is, define $q_{tax}(p_0)$ as the top 10% threshold in the tax (normalized so that the top 10% of the tax data has an average of s/p_0), and $q_{svy}(p_0)$ as the top 10% threshold in the survey (normalized so that the bottom 90% in the survey has an average of $(1-s)/(1-p_0)$). Define:

$$x_{AB}(p) = p + \frac{L(p) - L(p_0)}{q_{tax}(p_0)}$$
$$x_{AC}(p) = p + \frac{L(p) - L(p_0)}{q_{svy}(p_0)}$$

Then, for each value $pL(p)$, define a new percentile as:

$$p^* = \begin{cases} x_{AO}(p) + (x_{AB}(p) - x_{AO}(p))\frac{p - x_{AO}(p)}{x_{AC}(p) - x_{AO}(p)} & \text{if} \quad p < p_0 \\ p & \text{if} \quad p > p_0 \end{cases}$$

The new Lorenz curve $(p^*, L(p))$ defines a new tabulation, which has all the desired properties, and which can be interpolated again to retrieve the full set of g-percentiles.

7.3 Wealth

The measurement of wealth inequality in many countries remains a considerable challenge. For income, we can at least resort to surveys that, for the recent decades, are available in most countries. In our view, they remain poor proxies for the true levels of inequality. But at least they provide some information.

In comparison, the data sources for the measurement of the income distribution are often nonexistent, even in developed countries. Still, some progress has been made in recent years when it comes to survey data. In particular, at the initiative of the ECB, most European countries have started to regularly measure household wealth within the HFCS.

On the tax data side, little to no progress has been made — and in some cases the situation has actually regressed. The abandonment of wealth and sometimes inheritance taxes in several countries — as well as the separate taxation of capital flows — has made it increasingly difficult to track top wealth holders.

Because wealth is extremely concentrated, measuring it using surveys is even more complicated than income. Moreover, wealth surveys feature important methodological differences that make direct comparison of survey estimates difficult. The SCF in the United States is relatively good at oversampling the richest, thanks to important efforts by the Federal Reserve in this domain. The French or Spanish wealth surveys are also good examples. But many other countries perform little to no oversampling. As result, even within the ECB wealth survey — which is supposed be harmonized at the European level — differences between countries are hard to interpret. In any case, even strong forms of oversampling often appear insufficient, as exhibited by the fact that top capital incomes usually remain lower in wealth surveys that in tax data.

In the future, we insist on the need for countries to improve their statistical apparatus to track top wealth holders, by directly exchanging information with financial institutions and release anonymized data based on administrative wealth registers.

An auxiliary source of information that we can use while we wait for better data are "rich lists" such as the *Forbes Billionaires* list. These lists track the wealth of the richest individuals in a country, or in the case of Forbes the whole world. Many countries have magazines that publish such lists (for example the *Sunday Times* in the United Kingdom, *Challenges* in France, *Manager Magazin* in Germany). While they are an interesting source, they remain problematic and hard to exploit in a systematic way.

The first issue with these lists is that their methodology lacks transparency, and the concepts of wealth they rely on are not always well defined. The second issue is that they use inconsistent units of observations. While *Forbes* tracks the wealth of individuals (with some exceptions), many national lists group members of extended family into a single unit (which may in the worst cases include tens or hundreds of people). This makes wealth levels hard to compare. The third issue is that they only concern a tiny fraction of the population, so it is often hard to extrapolate the distribution over the entire distribution based on such limited data. This is especially the case in small or relatively poor countries that only feature a handful of billionaires at best.

In view of the situation, more research — and better data — is needed. Our methodological recommendations in this remain extremely open and emphasize the need to exploit whatever data is available. We discuss in Bajard, Chancel, et al., 2021 our methodology to produce wealth series for all countries, based on the sparse data at our disposal. These estimates are tentative and subject to revision (see also section 8.5).

Key Points

- In many countries, the data that is available is too limited to construct fully satisfying DINA series. We nonetheless make our best to cover these countries, making simple and transparent assumptions whenever necessary. We believe that this approach is better than the alternatives. Not to publish anything in these countries would be an excessively skeptical stance, since some data does exist, and is in fact widely used and discussed in academic and policy circles. To publish the data without adjustments would systematically bias comparisons — for example between countries with or without tax data, or countries with consumption rather income data. Nevertheless, we stress that these series are more fragile than the rest, and may subject to significant revisions as soon as better data becomes available.

- We start from whatever data is available, and apply simple corrections by g-percentile to harmonize the concepts measured in surveys (e.g., income vs. consumption), the statistical unit (e.g., households vs. equal-split adults), and the underrepresentation of the rich. In general, these correction profiles are inferred from neighboring countries.

- We incorporate information from tax data to our estimates whenever it is available, even if that information is very partial.

- We finally attribute missing income components to the various g-percentiles of the distribution to move from fiscal/survey income to national income. To that end, we use whatever data is at our disposal. But we provide default parameters to use as a last resort in the online appendix of these guidelines.

Chapter 8

Regional and Global Inequality

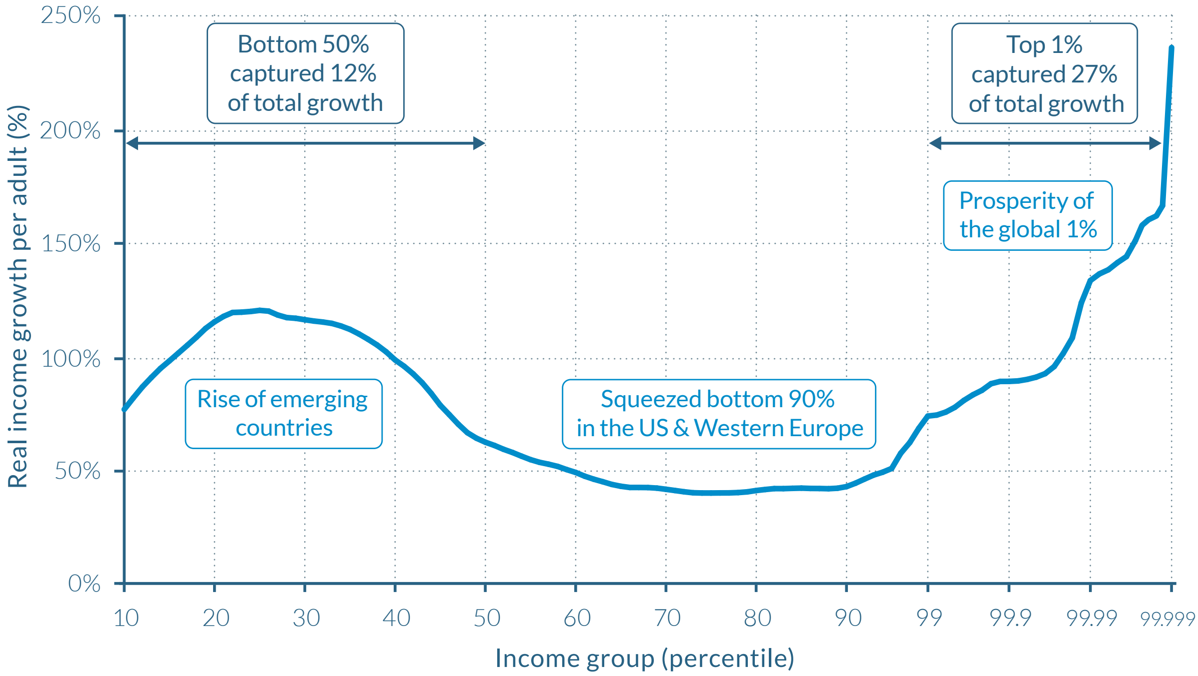
The dynamics of global inequality have attracted growing attention in recent years. Recent studies of global inequality combine household surveys and provide valuable estimates (Lakner and Milanovic, 2016; Liberati, 2015; Ortiz and Cummins, 2011). But as discussed above, surveys are not uniform across countries, they do not capture top incomes well, and are not consistent with macroeconomic totals.

By homogenizing different data sources in order to provide estimates of income and wealth inequality consistent with national accounts totals, the DINA methodology makes it possible to produce better systematic regional and global estimates of inequality. In this section, we describe the various methods followed to produce regional and global income and wealth distributional statistics.

8.1 Global Income Inequality Estimates

A first version of global income distribution consistent with national accounts estimates was published in the World Inequality Report 2018 (Alvaredo, Chancel, et al., 2018). In this exercise, authors start from national income aggregates and population series — available for all countries available on the WID, see section 4. Within each country, national income is the distributed to adult individuals (using the equal-split benchmark, see section 1.1.3) in countries with available DINA series. These countries represented around 2/3 of global income in late 2017 (see Chancel and Gethin, 2017, for methodological details).

The distribution of income in countries without DINA series as of late 2017 was estimated based



Source: World Inequality Report 2018 (Alvaredo, Chancel, et al., 2018). See wir2018.wid.world for more details. *Interpretation:* On the horizontal axis, the world population is divided into a hundred groups of equal population size and sorted in ascending order from left to right, according to each group's income level. The top 1% group is divided into ten groups, the richest of these groups are also divided into ten groups, and the very top group is again divided into ten groups of equal population size. The vertical axis shows the total income growth of an average individual in each group between 1980 and 2016. For percentile group p99p99.1 (the poorest 10% among the world's richest 1%), growth was 74% between 1980 and 2016. The top 1% captured 27% of total growth over this period. Income estimates account for differences in the cost of living between countries. Values are net of inflation.

Figure 8.1: Global Income Growth and Inequality, 1980–2016

on the distribution of income in neighboring countries for which they were available. More precisely, we followed two strategies. In strategy 1, we assume that, each year, countries without data had the same income distribution as their region's average ("static calibration"). In strategy 2, we assume that income growth was distributed in these countries in the same way as in their region's average ("dynamic calibration"). Given the share of global income and population covered by DINA series even at that time, estimates were relatively robust to the choice of the calibration strategy (see Chancel and Gethin, 2017; Clarke et al., 2017). However, these methods are obviously oversimplifications. More recent work in the context of the DINA project and the WID make it possible to improve inequality data for missing countries.

We use the merge function of the `gpinter` R package (see section 5.2.1) to aggregate distributions. It combines interpolated national distributions (also factoring in the population of each country) and generates 127 merged income g-percentiles. It can also produce a geographical breakdown of the population within each global g-percentile. One question is what currency conversion factor to apply between countries: PPPs or market exchange rates (see also section 1.2.2). There are good arguments in favor of aggregating national distributions using market exchange rates or PPPs. In a world where all goods and services are purchased at home, aggregating national income distributions using PPPs might seem like the best option in order to take into account price differentials between nations and avoid artificially inflating global income inequalities between individuals.

At the same time, an increasing number of services are bought by individuals outside their home countries, thanks to the internet and communication technologies as well as international travel. It may then be argued that the use of PPPs tends to reduce the true level of global inequality between individuals and that market exchange rates are more appropriate. While benchmark estimates of the World Inequality Report used PPP aggregates, we also published market exchange rates series in order to account for this complex reality.

8.2 Regional Income Inequality Updates

Following the World Inequality Report 2018, we started to perform new regional DINA estimates for world regions, based on the most up to date and available national accounts, survey and tax data.[1] They make it possible to revise global income distribution statistics published in the WID. These updated (or novel) national and regional series include:

- Europe DINA estimates (covering 38 European countries, excluding Belarus, Russia and Ukraine), by Blanchet, Chancel, and Gethin (2022).
- Africa DINA estimates (covering 55 countries), by Chancel, Cogneau, et al. (2019). These series are based on significantly less distributional information than the European ones (see chapter 7) but nonetheless represent an improvement compared to earlier estimates, because we attempt to express all series in per adult income, rather than relying on a mix of consumption, post-tax and pretax income data as was done in earlier regional or global aggregation work.
- Latin America DINA estimates (covering about 20 countries) by De Rosa, Flores, and Morgan (2024).

[1] Either publicly or through restrictions.

- Asia DINA estimates (covering about 40 countries) by Jenmana, Yang, and Khaled (2020).

Starting October 2020, the WID will include distributional statistics for each world region (and each country within these regions) up to year $n-1$. When there is no available new distributional information for a country between year $n-2$ and year $n-1$ (i.e., when there is no newly available tax or survey data to researchers), inequality levels are assumed to be constant while income levels and population evolve as per national accounts and population statistics. These statistics will be revised as soon as novel inequality data is released. This general strategy makes it possible to produce and update regional and global inequality statistics every year, without omitting countries because of a lack of data.

8.3 Adjustments for the Bottom of the Income Distribution

The lower deciles of income distributions (especially on a pretax basis) can exhibit negative values that we mostly consider undesirable. Negative incomes are possible in principle (for example, a business owner whose losses exceed their revenue). But in practice the treatment of these values is too heterogeneous across sources to get internationally comparable estimates. Moreover, negative incomes map poorly to our collective understanding of extreme poverty, so they should not be used for that purpose. In addition, they are large variations in the share of individuals with zero incomes at the bottom, which are due to methodological differences rather than reality, and we try to limit this phenomenon.

For these reasons, we choose to redistribute incomes within the bottom 20% in most countries, as a way to get more comparable estimates, using the following method. Unless the data is considered reliable enough not to warrant an adjustment, we bottom code incomes at zero, and impose that the bottom 5% have zero income.

The bottom of the distribution is divided in percentiles $p_0 = 0\%, p_1 = 1\%, \dots, p_{100} = 100\%$, and we wish to redistribute income within the bottom to avoid artificial methodological discrepancies.

We will consider that all incomes from p_0 to p_k are set to zero, and then we will redistribute incomes between p_{k+1} and p_n. So in our case we consider $k = 5$ and $n = 20$.

Let a_k be the average between percentile p_k and p_{k+1}. We assume that we know a_n, and also m, the average of the bracket $[p_{k+1}, p_n]$. We will assume the following form for the quantile function in that bracket:

$$Q(p) = a_n \left(\frac{p - p_{k+1}}{p_n - p_{k+1}} \right)^{\alpha}$$

which ensures that it starts at zero at p_{k+1} and then reaches a_n at p_n. We will choose α to match the mean of $[p_{k+1}, p_n]$, i.e.:

$$\int_{p_{k+1}}^{p_n} a_n \left(\frac{p - p_{k+1}}{p_n - p_{k+1}} \right)^{\alpha} \mathrm{d}p = m(p_n - p_{k+1})$$

Solving that equation leads to:

$$\alpha = \frac{a_n}{m} - 1$$

Then, the average between percentiles p_i and p_{i+1} is:

$$a_n \frac{[(p_{i+1} - p_{k+1})^{1+\alpha} - (p_i - p_{k+1})^{1+\alpha}]}{(1+\alpha)(p_n - p_{k+1})^{\alpha}(p_{i+1} - p_i)}$$

and the threshold of p_k is given by the formula of $Q(p)$.

8.4 Countries with No Income Inequality Data

A few countries do not produce any survey or tax data at all (e.g., Libya). For regional and global aggregation, we assume that these countries have the average inequality level of their region. Because these imputations are, in fact, used in our computations, we also report them in the WID for informational and transparency purposes. But we explicitly flag them as imputations, with a specific label and precaution note (see also chapter 9 on inequality data transparency indexes).

8.5 Global Wealth Aggregation

National-level official aggregate and distributional statistics are scarcer for wealth than for income. In the World Inequality Report 2018, we carried out a preliminary global wealth aggregation, focusing on the United States, Europe and China as a proxy for the global distribution of wealth (see Alvaredo, Chancel, et al., 2018, chapter 4). This work was carried forward in the World Inequality Report 2022, thanks to a better coverage of wealth aggregates data and preliminary country-level wealth inequality estimates for all countries. The method followed to produce these estimates is detailed in Bajard, Bauluz, et al. (2025).

8.6 Global Agricultural Land Inequality

The WIL, in collaboration with the International Land Coalition, has started a project on the measurement of the distribution of agricultural land in developing countries (Bauluz, Govind, and Novokmet, 2020). The goal of the project is to provide estimates of land inequality in as many countries as possible, based on the availability of data. This project departs from traditional estimates of land inequality that used agricultural land censuses (e.g. Deininger and Squire, 1998; Frankema, 2010), because this source faces various challenges.

First, land distribution calculated using the agricultural census captures the distribution of operational holdings (i.e., economic units of agricultural production under single management) rather than land ownership. From a distributional point of view, the latter is more relevant because agricultural census does not necessarily account for multiple landholdings per owner and fails to capture the full extent of land concentration. Second, differences in the value and quality of land are not measured in agriculture censuses, since they only report the land size of agricultural holdings. Additionally, census data, by definition, does not account for the landless households. This may not portray the actual levels of inequality or provide comparable estimates across countries.

To overcome these issues, the WIL's project exploits household surveys to provide estimates on land area and value distribution, as well as including the landless population in different countries across the world. There are two main types of surveys that are used in this project: World Bank's LSMS and official household surveys of different countries. The first two types of surveys generally comprise an agricultural module which collects information about the fields or plot owned by the household. The relevant information for estimating land ownership inequality is the land area, reported value and an indication of ownership.

The object of analysis is to measure the distribution of land ownership. Land ownership is defined as any agricultural land over which the household has private property rights. This is fairly consistently defined across countries. However, there are certain cases where private rights are less clearly defined. This is the case, for instance, in China and Vietnam, where rural household are provided extensive rights over the land despite the absence if strictly defined private ownership (e.g., rights to control, dispose and inherit the land Do and Iyer, 2003; Li and Zhao, 2008; McKinley and Griffin, 1993; Piketty, Yang, and Zucman, 2019). At the moment, this project does not include communal land in its definition of ownership. The plan is to investigate it in the future, as it plays a relevant role in certain countries (e.g., in Africa or Latin

America).

Two ways of measuring the agricultural land owned by a household are used. The first is in terms of area of agricultural land (i.e., the size of the land holdings owned by a household). The second way is in terms of value of agricultural land. The latter is the preferred measure since it accounts for the large heterogeneity of land types within a country and captures the value of land as an asset. Survey-reported values are based on the concept of current market value, where the agricultural land is valued at prevailing market prices.

Two population groups are used to measure land ownership inequality. The first population group is the landowners (i.e., those households owning land). The second group is the landowners plus the landless households. The latter should be seen as the benchmark unit, since it is important to account for the landless households to have a complete picture of the land inequality. Surveys are useful in examining the landless population, since they capture both the population of households living in rural areas as well as professional activities of each member of a household, including agriculture. This information, together with the number of households that are landowners, allows to identify the population of “landless households.” Landless households are defined as those where at least one of its members is employed in agriculture but does not report owning any agricultural land.

Key Points

- The first estimates of world inequality using DINA data were made and published for the World Inequality Report 2018 (Alvaredo, Chancel, et al., 2018).
- We intend to systematically combine or estimates of inequality by country to produce estimates of inequality at the regional and at the world level.
- We will update these estimates yearly based on the latest — and sometimes preliminary — information available.

Chapter 9

Data Quality and Transparency Measures

Accurate measurement of income and wealth inequality hinges on the release of detailed microdata, aggregated tabulations, and survey results by statistical agencies and governments. To make sense of the diverse information that becomes available, we assign quality grades to key distributional series and condense those grades into a single country-level score for a summarized view. Here we lay out, first, our upgraded grading framework (now with a time dimension and expanded to cover wealth and macroeconomic aggregates) and second, the construction and use of the Inequality Transparency Index.

9.1 Series Quality Grades

Key distributional series in the World Inequality Database (WID) are assigned a quality grade from 0 to 5, reflecting both the richness of their underlying data sources and the consistency of coverage over time. With this version of the guidelines, we have introduced an explicit time dimension to the grading framework—users can now track how data quality evolves year by year, spot improvements, identify gaps in coverage, and compare trends across countries. The detailed criteria for each grade are summarized in the tables 9.1 and 9.2. The first table lists the labels for each quality grade by series type; the second table provides the corresponding detailed descriptions.

These grades appear alongside each series in the WID database, giving a clear indication of both data coverage and time-series consistency.

Grade	Distributed income	Distributed wealth	Wealth aggregate	Income aggregate
0	Imputed regional average	Income-based weighting	Imputed regional average	Imputed regional average
1	Carry forward or backward	Adjusted surveys, no tax data	Extended from OLS prediction	Carry forward or backward
2	Interpolation	Simplified multiplier method	Predicted values (OLS)	Predicted values (OLS)
3	Surveys, but no tax data	National-accounts balancing	Extended from observed data (OLS)	Interpolation
4	Tax and survey data	Mixed capitalization–survey model	Extended from observed data (OLS)	Observed (old SNA or historical sources)
5	Tax and survey microdata	Wealth tax or administrative registers	Observed (latest SNA or academic paper)	Observed (latest SNA or academic paper)

Table 9.1: Quality Grade Labels

9.2 Improving Income and Wealth Inequality Data

A lot of effort is put by researchers associated to the DINA and WID projects to obtain the latest income and wealth micro data. Despite all these efforts, the current state of available distributional statistics remains far from satisfactory. In order to improve income and wealth distributional statistics worldwide, and absent micro data availability, basic informational tables should be produced by countries' tax or statistical authorities on an annual basis.

Such data tables are presented below (tables 9.3, 9.4 and 9.5). They include, for different income and wealth brackets, the number of individuals in each bracket as well as the decomposition of total income and total wealth by different types (labor income, capital income, financial wealth, non-financial wealth, etc.) Taxes paid by each bracket is also key in order to properly monitor the evolution of pretax vs. post-tax inequality estimates as well as tax incidence.

Net wealth refers to total assets (real estate, business, financial, etc.), net of debt. For country residents, all domestic and foreign assets should be included. For non-residents, all domestic assets should be included (in particular real estate assets located in the country, as well as all financial assets related to firms and economic activities conducted in the country). To the extent possible, their foreign assets should also be included.

Grade	Distributed income	Distributed wealth	Wealth aggregate	Income aggregate
0	No data is available (imputed regional average). This is the lowest-quality category and is used only as a last resort to ensure coverage.	Estimates are based on the observed correlation between income inequality and wealth inequality. Wealth distribution is predicted using a weighted average of other countries' distributions, with adjustments for rich-list and national accounts.	No data is available (imputed regional average). This is the lowest-quality category and is used only as a last resort to ensure coverage.	No data is available (imputed regional average). This is the lowest-quality category and is used only as a last resort to ensure coverage.
1	The value is imputed by carrying the nearest available observation forward or backward, filling short gaps in an otherwise continuous series.	Based solely on survey data, statistically adjusted to approximate missing top-end values; includes rich-list and national-accounts corrections.	Observed data are fitted to an OLS model with regional covariates; missing periods follow the regional average trajectory.	The value is imputed by carrying the nearest available observation forward or backward, filling short gaps in an otherwise continuous series.
2	Values are estimated by linear interpolation between two known data points, typically for short, internal gaps.	Wealth is imputed via a simplified mortality-based multiplier method, supplemented by partial microdata and rich-list/national-accounts adjustments.	Series are predicted using OLS regressions on country-specific covariates rather than past observations.	Values are estimated using OLS regressions on relevant macroeconomic covariates.
3	Only income or consumption survey data are used; no tax information is incorporated.	Aggregate wealth figures are derived from national-accounts balancing items, then distributed according to survey shares.	Earlier observed data exist; missing values are extrapolated via OLS when country-specific methods are infeasible.	Values are estimated by linear interpolation between two known data points, filling short, internal gaps.
4	Tax and survey data are both available, but at least one source is in tabulated (not micro) form.	Wealth distribution is modeled using a hybrid of income-capitalization techniques and survey information, with rich-list/national-accounts adjustments.	Observed data exist but with intermittent gaps; OLS extrapolation fills these gaps.	Values are directly sourced from older SNA frameworks or historical compilations; comparability with current standards may be limited.
5	Fully integrated micro-level data from tax records and household surveys are available.	High-quality wealth microdata from administrative registers or tax declarations, requiring only minor imputations for unobserved assets.	Directly observed in official balance sheets, global asset datasets, or validated academic studies.	Directly observed from the most recent national accounts under the latest SNA or benchmark academic research, representing the highest-quality data.

Table 9.2: Quality Grade Descriptions

net income bracket ($)	**number of individuals**	**total income**	**labor income**	**capital income**	*incl. housing asset income*	*incl. equity asset and net interest income*	*incl. pension and life insurance asset income*	**total income taxes**	*incl. personal income tax*	*incl. corporate income tax*	*incl. capital gains tax*	**total wealth taxes**	*incl. wealth and property tax*	*incl. inheritance and estate tax*
0–10k	...	...	...	...	...	...	...	...	...	...	...	...	...	...
10k–20k	...	...	...	...	...	...	...	...	...	...	...	...	...	...
20k–30k	...	...	...	...	...	...	...	...	...	...	...	...	...	...
30k–40k	...	...	...	...	...	...	...	...	...	...	...	...	...	...
40k–50k	...	...	...	...	...	...	...	...	...	...	...	...	...	...
50k–70k	...	...	...	...	...	...	...	...	...	...	...	...	...	...
70k–100k	...	...	...	...	...	...	...	...	...	...	...	...	...	...
100k–150k	...	...	...	...	...	...	...	...	...	...	...	...	...	...
150k–200k	...	...	...	...	...	...	...	...	...	...	...	...	...	...
200k–400k	...	...	...	...	...	...	...	...	...	...	...	...	...	...
400k–600k	...	...	...	...	...	...	...	...	...	...	...	...	...	...
600k–800k	...	...	...	...	...	...	...	...	...	...	...	...	...	...
800k–1m	...	...	...	...	...	...	...	...	...	...	...	...	...	...
1m–10m	...	...	...	...	...	...	...	...	...	...	...	...	...	...
10m–100m	...	...	...	...	...	...	...	...	...	...	...	...	...	...
>100m	...	...	...	...	...	...	...	...	...	...	...	...	...	...

Table 9.3: Data Table to Be Published by Tax Authorities: Data by Income Bracket

net wealth bracket ($)	**number if individuals**	*incl. number of residents*	*incl. number of non-residents*	**total net wealth**	*incl. residents*	*incl. non-residents*	**total wealth taxes**	*incl. wealth and property tax*	*incl. capital gains tax*	*incl. inheritance and estate tax*	**total income taxes**	*incl. personal income tax*	*incl. corporate income tax*
<0	...	...	...	...	...	...	...	...	...	...	...	...	...
0–10k	...	...	...	...	...	...	...	...	...	...	...	...	...
10k–100k	...	...	...	...	...	...	...	...	...	...	...	...	...
100k–1m	...	...	...	...	...	...	...	...	...	...	...	...	...
1m–10m	...	...	...	...	...	...	...	...	...	...	...	...	...
10m–100m	...	...	...	...	...	...	...	...	...	...	...	...	...
100m–1bn	...	...	...	...	...	...	...	...	...	...	...	...	...
1bn–5bn	...	...	...	...	...	...	...	...	...	...	...	...	...
5bn–10bn	...	...	...	...	...	...	...	...	...	...	...	...	...
>10bn	...	...	...	...	...	...	...	...	...	...	...	...	...

Table 9.4: Data Table to Be Published by Tax Authorities: Data by Wealth Bracket

net wealth bracket ($)	**number of indi-viduals**	**total wealth**	*incl. currency and deposits*	*incl. bonds and loans*	*incl. equities and fun shares*	*incl. pension funds and life insur-ance*	*incl. real estate*	*incl. business and other non-financial assets*	*incl. debt*	*incl. total domestic assets*	*incl. total foreign assets*	**total income**	*incl. capital income*	*incl. labor income*
<0	…	…	…	…	…	…	…	…	…	…	…	…	…	…
0–10k	…	…	…	…	…	…	…	…	…	…	…	…	…	…
10k–100k	…	…	…	…	…	…	…	…	…	…	…	…	…	…
100k–1m	…	…	…	…	…	…	…	…	…	…	…	…	…	…
1m–10m	…	…	…	…	…	…	…	…	…	…	…	…	…	…
10m–100m	…	…	…	…	…	…	…	…	…	…	…	…	…	…
100m–1bn	…	…	…	…	…	…	…	…	…	…	…	…	…	…
1bn–5bn	…	…	…	…	…	…	…	…	…	…	…	…	…	…
5bn–10bn	…	…	…	…	…	…	…	…	…	…	…	…	…	…
>10bn	…	…	…	…	…	…	…	…	…	…	…	…	…	…

Table 9.5: Data Table to Be Published by Tax Authorities: Wealth and Income Composition by Wealth Bracket

Key Points

- Given the heterogeneity in the quality of the data available worldwide, the quality of the estimates available on the WID varies a lot.
- We make our best to communicate on the quality of each series by giving them a simple note — from 0 to 5 — based on broad considerations about the data that they use. We make this ranking directly visible on our website. For precise details, we refer users to country and region-specific papers.
- To improve the quality of our series, many governments will have to publish better data on the distribution of both income and wealth. To encourage them to do so, we publish a country-specific data transparency index that measures the amount and the quality of the data made available to researchers.
- We give examples of the type of tax tabulations that — in our view — constitutes the bare minimum of what every tax administration or national statistical institute in the world should be publishing. Whenever possible, government should go further and give researchers direct access to tax microdata.

List of Tables

List of Figures

Acronyms

AMA Analysis of Main Aggregates. 90–92

BOPS Balance of Payments Statistics. 100, 101

BPM6 Sixth Edition of the Balance of Payments and International Investment Position Manual. 20

CFC consumption of fixed capital. 40, 43, 95, 96, 99, 111

CPI consumer price index. 25, 26, 92

CPIS Coordinated Portfolio Investment Survey. 101

CPS Current Population Survey. 130

DINA Distributional National Accounts. 11, 13, 15, 20, 21, 23, 24, 32, 36–39, 45, 50, 51, 57, 58, 61, 63, 72, 78, 86, 89, 90, 114, 133, 136, 156–159, 166–169, 174, 176

ECB European Central Bank. 164, 165

ERC European Research Council. 14

ERFS Enquête revenus fiscaux et sociaux. 130

ESA European System of Accounts. 38, 75

EU-SILC European Union Statistics on Income and Living Conditions. 124, 128

Eurostat European Statistical Office. 51

FDI foreign direct investment. 41, 100

GDP gross domestic product. 25–27, 29, 36, 38–40, 43, 57, 62, 77, 90–93, 96, 100, 101

HFCS Household Finance and Consumption Survey. 164

ICP international comparison program. 29, 36

IMF International Monetary Fund. 20, 48, 91–93, 99–101

LCU local currency unit. 25, 30

LSMS Living Standards Measurement Surveys. 172

MADT Main Aggregates and Detailed Tables. 91–93

MICS Mixed Income Capitalization-Survey. 137, 148

NFI net foreign income. 40

NNI net national income. 39, 40, 43

NPISH non-profit institutions serving households. 51, 54, 61, 80, 86

OECD Organisation for Economic Co-operation and Development. 29, 48, 51, 92, 93, 99–101, 158

Pacs Pacte civil de solidarité. 22

PCE personal consumption expenditures price index. 25

PPP purchasing power parity. 25, 27–31, 36, 96, 169

PSE Paris School of Economics. 11

SCF Survey of Consumer Finances. 140, 165

SNA System of National Accounts. 11, 13, 15, 20, 23, 38–49, 51–54, 57, 59, 60, 65, 66, 72–79, 81–83, 91, 104, 126–128, 130–133, 136

UN United Nations. 90–93

UNU-WIDER United Nations University World Institute for Development Economics Research. 150

VAT value-added tax. 53, 59, 62, 63, 72, 129

WEO World Economic Outlook. 91, 92

WID World Inequality Database. 11–15, 19, 23, 25, 27, 28, 30, 31, 36, 38, 39, 43–45, 51, 54, 55, 73, 75, 76, 78, 89, 93, 102–106, 111, 127, 142, 143, 145, 149, 167–171, 176, 181

WIID World Income Inequality Database. 150

WIL World Inequality Lab. 11, 70, 90–92, 100, 101, 157, 172

WPP World Population Prospects. 90

WTID World Top Incomes Database. 12, 13, 19, 20, 37, 155, 160

Index

Bibliography

Albers, T. N. H., C. Bartels, and M. Schularick (2025). "Wealth and its Distribution in Germany, 1895-2018". In: *Journal of the European Economic Association, forthcoming* (cit. on p. 104).

Allen, C. (2019). "Essays in International Macroeconomics". PhD thesis. Trinity College Dublin (cit. on p. 99).

Alstadsæter, A., M. Jacob, W. Kopczuk, and K. Telle (2017). "Accounting for Business Income in Measuring Top Income Shares: Integrated Accrual Approach Using Individual and Firm Data from Norway". In: NBER Working Paper Series 22888 (cit. on pp. 58, 127).

Alstadsæter, A., N. Johannesen, and G. Zucman (2018). "Who owns the wealth in tax havens? Macro evidence and implications for global inequality". In: *Journal of Public Economics* 162, pp. 89–100 (cit. on p. 44).

Alvaredo, F., L. Assouad, and T. Piketty (2019). "Measuring Inequality in the Middle East 1990–2016: The World's Most Unequal Region?" In: *Review of Income and Wealth* 65.4, pp. 685–711 (cit. on p. 156).

Alvaredo, F., A. B. Atkinson, and S. Morelli (2018). "Top wealth shares in the UK over more than a century". In: *Journal of Public Economics* 162, pp. 26–47 (cit. on pp. 15, 140).

Alvaredo, F., A. B. Atkinson, T. Piketty, and E. Saez (2013). "The top 1 percent in international and historical perspective". In: *Journal of Economic Perspectives* 27.3, pp. 3–20 (cit. on p. 12).

Alvaredo, F. and A. B. Atkinson (2010). "Colonial Rule, Apartheid and Natural Resources: Top Incomes in South Africa, 1903-2007". In: Centre for Economic Policy Research Working Paper Series 8155 (cit. on p. 152).

Alvaredo, F., L. Chancel, T. Piketty, E. Saez, and G. Zucman (2018). World Inequality Report. World Inequality Lab (cit. on pp. 167, 168, 171, 174).

Artola Blanco, M., L. Bauluz, and C. Martínez-Toledano (2021). "Wealth in Spain 1900–2017 a Country of two Lands". In: *The Economic Journal* 131.633, pp. 129–155 (cit. on p. 104).

Atkinson, A. B. (2005). “Comparing the Distribution of Top Incomes Across Countries”. In: *Journal of the European Economic Association* 3.2-3, pp. 393–401 (cit. on p. 20).
– (2007). “Measuring Top Incomes: Methodological Issues”. In: *Top Incomes over the Twentieth Century: A Contrast Between Continental European and English-Speaking Countries* (cit. on p. 20).
Atkinson, A. B. and A. J. Harrison (1978). Distribution of personal wealth in Britain. Cambridge University Press (cit. on pp. 12, 14, 137, 138, 142).
Atkinson, A. B. and T. Piketty (2007). Top incomes over the twentieth century: a contrast between continental European and English-speaking countries. Oxford University Press (cit. on p. 12).
– (2010). Top incomes: a global perspective. Oxford University Press (cit. on p. 12).
Atkinson, A. B., T. Piketty, and E. Saez (2011). “Top incomes in the long run of history”. In: *Journal of Economic Literature* 49.1, pp. 3–71 (cit. on p. 12).
Bach, S., C. Bartels, and T. Neef (2020). “Distributional National Accounts: A Macro-Micro Approach to Inequality in Germany” (cit. on pp. 114, 115, 137).
Bachas, P., M. H. Fisher-Post, A. Jensen, and G. Zucman (2022). “Globalization and Factor Income Taxation”. In: WID.world Working Paper Series 2022/05 (cit. on p. 102).
Bachas, P., L. Gadenne, and A. Jensen (2024). “Informality, Consumption Taxes, and Redistribution”. In: *Review of Economic Studies* 91.5, pp. 2604–2634 (cit. on p. 129).
Bajard, F., L. Chancel, M. Rowaida, and P. Thomas (2021). “Global wealth inequality on WID.world: estimates and imputations”. In: WID.world Technical Note Series 2021/16 (cit. on p. 165).
Bajard, F., L. Bauluz, P. Brassac, L. Chancel, C. Martínez-Toledano, T. Piketty, and A. Sodano (2025). “Global wealth inequality on WID.world: Estimates and imputations”. In: *World Inequality Lab* (cit. on pp. 15, 147, 171).
Barnett, G. E. (1936). Two Tracts by Gregory King. Johns Hopkins University Press (cit. on p. 14).
Baselgia, E. and I. Z. Martínez (2023). “Wealth-income ratios in free market capitalism: Switzerland, 1900–2020”. In: *Review of Economics and Statistics*, pp. 1–52 (cit. on p. 104).
Bauluz, L. (2017). “Revised national income and wealth series: Australia, Canada, France, Germany, Italy, Japan, UK and USA”. In: WID.world Working Paper 2017/23 (cit. on p. 104).
Bauluz, L., P. Brassac, C. Martínez-Toledano, T. Piketty, and A. Sodano (2024). “Estimation of Global Wealth Aggregates in WID.world: Methodology”. In: WID.world Technical Note Series 2024/14 (cit. on pp. 15, 103, 106, 107).

Bauluz, L., Y. Govind, and F. Novokmet (2020). “Global Land Inequality”. In: WID.world Working Paper Series 2020/10 (cit. on p. 172).
Blanchet, T., L. Chancel, and A. Gethin (2022). “Why is Europe less Unequal than the United States?” In: *American Economic Journal: Applied Economics* 14.4, pp. 480–518 (cit. on pp. 15, 23, 48, 67, 127, 129, 155, 158, 169).
Blanchet, T., I. Flores, and M. Morgan (2019). “The Weight of the Rich: Improving Surveys Using Tax Data”. In: WID.world Working Paper Series 2018/2 (cit. on pp. 115, 118–120, 122, 123, 136, 152, 159–161).
Blanchet, T., J. Fournier, and T. Piketty (2017). “Generalized Pareto Curves: Theory and Applications”. In: WID.world Working Paper Series 2017/3 (cit. on pp. 116, 150, 151, 160).
– (2022). “Generalized Pareto curves: theory and applications”. In: *Review of Income and Wealth* 68.1, pp. 263–288 (cit. on p. 146).
Bozio, A., B. Garbinti, J. Goupille-Lebret, M. Guillot, and T. Piketty (2020). Predistribution versus Redistribution: Evidence from France and the United States. Working Paper 2020/22. World Inequality Lab (cit. on pp. 13, 15, 114, 129).
Bricker, J., A. Henriques, and P. Hansen (2018). “How much has wealth concentration grown in the United States? A re-examination of data from 2001-2013”. In: Finance and Economics Discussion Series (cit. on p. 139).
Bruil, A., C. van Essen, W. Leenders, A. Lejour, J. Möhlmann, and S. Rabaté (2025). “Inequality and Redistribution in the Netherlands”. Working Paper (cit. on pp. 69, 130).
Canberra Group (2011). Canberra Group Handbook on Household Income Statistics. Second Edition (cit. on p. 128).
Carranza, R., M. De Rosa, and I. Flores (2025). “Wealth inequality in Latin America (2000–2020): data, facts and conjectures”. In: *Oxford Open Economics* 4.Supplement_1 (cit. on p. 104).
Chancel, L., D. Cogneau, A. Gethin, and A. Myczkowski (2019). “How Large Are African Inequalities ? Towards Distributional National Accounts in Africa , 1990-2017”. In: WID.world Working Paper Series 2019/13 (cit. on pp. 15, 151, 152, 155, 169).
Chancel, L. and A. Gethin (2017). “Building a global income distribution brick by brick”. In: WID.world Technical Note Series 2017/5 (cit. on pp. 167, 168).
Chancel, L. and T. Piketty (Oct. 2021). “Global Income Inequality, 1820–2020: the Persistence and Mutation of Extreme Inequality”. In: *Journal of the European Economic Association* 19.6, pp. 3025–3062 (cit. on p. 15).

Chatterjee, A., L. Czajka, and A. Gethin (2020). Estimating the distribution of household wealth in South Africa. United Nations University World Institute for Development Economics Research (cit. on p. 104).

Chilès, J.-P. and N. Desassis (2018). "Fifty Years of Kriging". In: *Handbook of Mathematical Geosciences: Fifty Years of IAMG*. Ed. by B. Daya Sagar, Q. Cheng, and F. Agterberg. Cham: Springer International Publishing, pp. 589–612 (cit. on p. 96).

Clarke, R., L. Chancel, R. Clarke, and A. Gethin (2017). "World Inequality Report 2018: Technical Notes for Figures and Tables". In: WID.world Technical Note Series 2017/8 (cit. on p. 168).

Conte, M., P. Cotterlaz, T. Mayer, et al. (2022). The CEPII gravity database. CEPII (cit. on p. 108).

Currier, L., E. L. Glaeser, and G. E. Kreindler (2024). "Infrastructure Inequality: Who Pays the Cost of Road Roughness?" Working Paper (cit. on p. 70).

Czajka, L. (2017). "Income Inequality in Côte d'Ivoire: 1985-2014". In: *WID.world Working Paper Series* 2017/8 (cit. on pp. 123, 152).

Daly, P. and M. Morgan (2021). "The Transformation of Wealth in a Small Open Economy: the Case of Ireland, 1995-2019". In: *Working Paper* (cit. on p. 104).

Davies, J. B., R. Lluberas, and A. F. Shorrocks (2017). "Estimating the level and distribution of global wealth, 2000–2014". In: *Review of Income and Wealth* 63.4, pp. 731–759 (cit. on p. 105).

Davies, J. B., S. Sandström, A. Shorrocks, and E. N. Wolff (2011). "The level and distribution of global household wealth". In: *The Economic Journal* 121.551, pp. 223–254 (cit. on p. 105).

De Rosa, M., I. Flores, and M. Morgan (2024). "More unequal or not as rich? Revisiting the Latin American exception". In: *World Development* 184, p. 106737 (cit. on pp. 15, 114, 169).

Deininger, K. and L. Squire (1998). "New ways of looking at old issues: Inequality and growth". In: *Journal of Development Economics* 57.2, pp. 259–287 (cit. on p. 172).

Deville, J.-C. (2000). "Generalized calibration and application to weighting for non-response". In: *COMPSTAT: Proceedings in Computational Statistics 14th Symposium held in Utrecht, The Netherlands, 2000*. Ed. by J. G. Bethlehem and P. G. M. van der Heijden. Heidelberg: Physica-Verlag HD, pp. 65–76 (cit. on p. 123).

Dietrich, J., G. Nievas, M. Odersky, T. Piketty, and A. Somanchi (2025). "Extending WID National Accounts Series: Institutional Sectors and Factor Shares". In: WID.world Technical Note Series 2025/03 (cit. on pp. 56, 102).

Do, Q.-t. and L. Iyer (2003). "Land Rights and Economic Development :" (cit. on p. 172).

ECB (2019). “Developing distributional financial accounts – Final report of the Expert Group on Linking Macro and Micro data for the Household Sector” (cit. on p. 14).

Feenstra, R. C., R. Inklaar, and M. P. Timmer (2015). “The Next Generation of the Penn World Table”. In: *American Economic Review* 105.10, pp. 3150–3182 (cit. on p. 29).

Fisher-Post, M. and A. Gethin (2023). “Government Redistribution and Development: Global Estimates of Tax-and-Transfer Progressivity, 1980-2019”. Working Paper (cit. on pp. 70, 128, 129).

Frankema, E. (2010). “The colonial roots of land inequality: Geography, factor endowments, or institutions?” In: *Economic History Review* 63.2, pp. 418–451 (cit. on p. 172).

Gabrielli, V., T. Neef, and A.-S. Robilliard (2024). “2024 Update for Female Labor Income Share”. In: WID.world Technical Note Series 2024/13 (cit. on p. 134).

Garbinti, B., J. Goupille-Lebret, and T. Piketty (2018). “Income inequality in France, 1900–2014: Evidence from Distributional National Accounts (DINA)”. In: *Journal of Public Economics* 162, pp. 63–77 (cit. on pp. 13, 15, 67, 114, 125, 155, 157).

– (2020). “Accounting for Wealth-Inequality Dynamics: Methods, Estimates, and Simulations for France”. In: *Journal of the European Economic Association* (cit. on pp. 13, 15, 80, 137–141).

Germain, J.-M., M. André, and T. Blanchet (2021). “Rapport du groupe d’experts sur la mesure des inégalités et de la redistribution”. Insee Méthodes n. 138 (cit. on pp. 69, 130).

Gethin, A. (2024a). “A New Database of General Government Revenue and Expenditure by Function, 1980-2022”. WID Technical Note (cit. on p. 68).

– (2024b). “Revisiting Global Poverty Reduction: Public Goods and the World Distribution of Income, 1980-2022”. Working Paper (cit. on p. 70).

– (2024c). “Who Benefits from Public Services? Novel Evidence and Implications for Inequality Measurement”. Working Paper (cit. on pp. 69, 70, 130).

Gomez-Carrera, R., R. Moshrif, G. Nievas, T. Piketty, and A. Somanchi (2025). WID Population Series Extension: Projections 2024-2100 with Age and Gender Breakdowns. Technical Note 2024/12. World Inequality Lab, Paris School of Economics (cit. on pp. 15, 91).

Hodgson, G. M. (2014). “What is capital? Economists and sociologists have changed its meaning: Should it be changed back?” In: *Cambridge Journal of Economics* 38.5, pp. 1063–1086 (cit. on p. 77).

Hosking, J. R. M. and J. R. Wallis (1987). “Parameter and Quantile Estimation for the Generalized Pareto Distribution”. In: *Technometrics* 29.3, pp. 339–349 (cit. on p. 159).

IMF (2009). Balance of Payments and International Investment Positions, Sixth Edition. Washington, D.C. (cit. on p. 20).

Jenmana, T. (2018). “Democratisation and the Emergence of Class: Conflicts Income Inequality in Thailand” (cit. on p. 155).

Jenmana, T., L. Yang, and R. Khaled (2020). “Distributional National Accounts in Asia” (cit. on pp. 15, 170).

Kopczuk, W. (2015). “What Do We Know about the Evolution of Top Wealth Shares in the United States?” In: *Journal of Economic Perspectives* 29.1, pp. 47–66 (cit. on p. 139).

Kumar, R. (2019). “The evolution of wealth-income ratios in India 1860-2012”. In: *Available at SSRN 3111846* (cit. on p. 104).

Kuznets, S. (1953). Shares of Upper Income Groups in Income and Savings. National Bureau of Economic Research (cit. on pp. 12, 14).

Lakner, C. and B. Milanovic (2016). “Global Income Distribution: From the Fall of the Berlin Wall to the Great Recession”. In: *The World Bank Economic Review* 30.2, pp. 203–232 (cit. on p. 167).

Lane, P. R. and G. M. Milesi-Ferretti (2018). “The External Wealth of Nations Revisited: International Financial Integration in the Aftermath of the Global Financial Crisis”. In: *IMF Economic Review* 66.1, pp. 189–222 (cit. on pp. 100, 101, 108).

Lequiller, F. and D. Blades (2014). Understanding National Accounts: Second Edition (cit. on pp. 26, 39, 43, 47).

Li, S. and R. Zhao (2008). “Changes in the Distribution of Wealth in China, 1995–2002”. In: *Personal Wealth from a Global Perspective*. Oxford: Oxford University Press (cit. on p. 172).

Liberati, P. (2015). “The world distribution of income and its inequality, 1970-2009”. In: *Review of Income and Wealth* 61.2, pp. 248–273 (cit. on p. 167).

Lustig, N. (2018). Commitment to Equity Handbook: Estimating the Impact of Fiscal Policy on Inequality and Poverty. Brookings Institution Press (cit. on p. 70).

Maddison, A. (2007). Contours of the World Economy 1-2030 AD. Oxford University Press (cit. on pp. 91, 92).

– (2010). “Background Note on "Historical Statistics"” (cit. on p. 29).

Maddison, A. and H. X. Wu (2007). “China’s Economic Performance: How Fast has GDP Grown; How Big is it Compared with the USA?” (Cit. on p. 92).

Martínez-Toledano, C. (2020). “House Price Cycles, Wealth Inequality and Portfolio Reshuffling” (cit. on pp. 137, 139).

McCarthy, P. (2013). “Extrapolating PPPs and Comparing ICP Benchmark Results”. In: *Measuring the Real Size of the World Economy: The Framework, Methodology, and Results of the International Comparison Program*, pp. 441–472 (cit. on p. 29).

McKinley, T. and K. Griffin (1993). “The distribution of land in rural China”. In: *The Journal of Peasant Studies* 21.1, pp. 71–84 (cit. on p. 172).

Mo, Z., Q. Wang, L. Yang, and X. Zhang (2025). “Ideological Shifts and Wealth Dynamics: Unraveling the Century-Long Accumulation of Chinese National Wealth (1911-2020)”. In: (cit. on p. 104).

Morgan, M. (2017). “Falling Inequality beneath Extreme and Persistent Concentration: New Evidence for Brazil Combining National Accounts, Surveys and Fiscal Data, 2001-2015”. In: WID.world Working Paper Series 2017/12 (cit. on p. 155).

Neef, T. and A.-S. Robilliard (2021). “Half the Sky? The Female Labor Income Share in a Global Perspective”. In: WID.world Working Paper Series 2021/22 (cit. on p. 70).

Nievas, G. and T. Piketty (2025a). “Unequal exchange and North–South relations: Evidence from global trade flows and the world balance of payments 1800–202”. In: (cit. on p. 109).

– (May 2025b). WID National Accounts Series: Updated and Extended Coverage 1800–2023. Technical Note 2025/02. Version of May 28, 2025. Paris, France: World Inequality Lab, Paris School of Economics (cit. on pp. 15, 33, 109).

Nievas, G. and T. Piketty (2024). “Allocating Missing Foreign Wealth And Missing Foreign Income: Sources And Methods Used In Wid. World”. In: (cit. on pp. 15, 44, 85, 108).

Nievas, G. and A. Sodano (2024). “Has the US Exorbitant Privilege Become a Rich World Privilege? Rates of Return and Foreign Assets from a Global Perspective, 1970-2022”. In: (cit. on pp. 44, 104–107).

Novokmet, F. and N. Kump (2018). “Top incomes in Croatia and Slovenia, from 1960s until today” (cit. on p. 115).

Novokmet, F., T. Piketty, and G. Zucman (2018). “From Soviets to oligarchs: inequality and property in Russia 1905–2016”. In: *The Journal of Economic Inequality* 16.2, pp. 189–223 (cit. on p. 104).

OECD (2019). “Distributional information on Household income, consumption and saving – Guidelines” (cit. on p. 14).

– (2020). “Handbook on Compiling Distributional Results on Household Income, Consumption and Saving Consistent with National Accounts” (cit. on p. 14).

Ortiz, I. and M. Cummins (2011). “Global Inequality: Beyond the Bottom Billion – A Rapid Review of Income Distribution in 141 Countries” (cit. on p. 167).

Petersen, K. B. and M. S. Pedersen (2012). “The Matrix Cookbook” (cit. on p. 97).

Piketty, T. (2001). Les hauts revenus en France au XXe siècle. Inégalités et redistributions 1901-1998. Grasset (cit. on pp. 11, 12, 19).

Piketty, T. (2003). “Income Inequality in France, 1901–1998”. In: *Journal of Political Economy* 111.5, pp. 1004–1042 (cit. on pp. 11, 12, 19).

– (2011). “On the long-run evolution of inheritance: France 1820-2050”. In: *Quarterly Journal of Economics* 126.3, pp. 1071–1131 (cit. on p. 140).

– (2014). Capital in the Twenty First Century. Cambridge, Massachusetts: The Belknap Press of Harvard University Press (cit. on pp. 12, 75, 84).

Piketty, T. and E. Saez (2003). “Income Inequality in the United States, 1913–1998”. In: *Quarterly Journal of Economics* CXVIII.1 (cit. on pp. 11, 12, 19).

Piketty, T., E. Saez, and G. Zucman (May 2018). “Distributional National Accounts: Methods and Estimates for the United States”. In: *The Quarterly Journal of Economics* 133.2, pp. 553–609 (cit. on pp. 13, 15, 64, 67, 69, 76, 114, 125, 129, 157).

– (2019). “Simplified Distributional National Accounts”. In: *AEA Papers and Proceedings* 109, pp. 289–295 (cit. on p. 157).

Piketty, T., L. Yang, and G. Zucman (2019). “Capital Accumulation, Private Property, and Rising Inequality in China, 1978–2015”. In: *American Economic Review* 109.7, pp. 2469–96 (cit. on pp. 13, 15, 23, 172).

Piketty, T. and G. Zucman (2014). “Capital is Back: Wealth-Income Rations in Rich Countries 1700–2010”. In: *Quarterly Journal of Economics* 129.3, pp. 1255–1310 (cit. on pp. 12, 26, 54, 74, 75, 83, 84, 104, 106).

Ponthieux, S. and D. Meurs (2015). “Gender inequality”. In: *Handbook of income distribution*. Vol. 2. Elsevier, pp. 981–1146 (cit. on p. 133).

R Core Team (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria.

Riedel, L. and H. Stichnoth (2022). “Government expenditure in the DINA framework: Allocation methods and consequences for post-tax income inequality”. Working Paper (cit. on p. 69).

Robbins, J. A. (2018). “Capital Gains and the Distribution of Income on the United States” (cit. on p. 58).

Saez, E. and G. Zucman (Feb. 2016). “ Wealth Inequality in the United States since 1913: Evidence from Capitalized Income Tax Data”. In: *The Quarterly Journal of Economics* 131.2, pp. 519–578 (cit. on pp. 13, 76, 137, 138, 140, 142, 143).

– (2019). “Clarifying Distributional Tax Incidence: Who Pays Current Taxes vs. Tax Reform Analysis” (cit. on p. 61).

Törmälehto, V.-M. (2019). “Reconciliation of EU statistics on income and living conditions (EU-SILC) data and national accounts” (cit. on p. 14).

Toussaint, S. J., A. de Vicq, M. Moatsos, and T. van der Valk (2025). “Wealth-Income Ratios in a Small, Open Economy: The Netherlands, 1854-–2019”. In: *European Economic Review* 178, p. 105099 (cit. on p. 104).

United Nations (2009). The System of National Accounts. New York (cit. on pp. 39, 45, 46, 49, 51–54, 73, 74, 77–79, 81–83).

Waldenström, D. (2017). “Wealth-Income Ratios in a Small, Developing Economy: Sweden, 1810–2014”. In: *The Journal of Economic History* 77.1, pp. 285–313 (cit. on p. 104).

Zucman, G. (2013). “The Missing Wealth of Nations, Are Europe and the U.S. net Debtors or net Creditors?” In: *Quarterly Journal of Economics* 128.3, pp. 1321–1364 (cit. on pp. 43, 85).

Zwijnenburg, J., S. Bournot, and F. Giovannelli (2017). “Expert Group on Disparities within a National Accounts framework – Results from a 2015 exercise” (cit. on p. 14).