



**Income Inequality in France, 1900-2014:  
Evidence from Distributional National Accounts (DINA)  
Data Appendix**

Bertrand Garbinti, Jonathan Goupille-Lebret, Thomas Piketty \*

First Version: July 12, 2016

Last Revised: April 10, 2017

This data appendix provides methodological details and complete data series for our paper “Income Inequality in France, 1900-2014: Evidence from Distributional National Accounts (DINA)”. It is supplemented by a set of data files and computer codes (GGP2017DINA.zip).

\*We are grateful to Facundo Alvaredo, Emmanuel Saez, Daniel Waldenström and Gabriel Zucman for numerous conversations. The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme, ERC Grant Agreement n. 340831. This work is also supported by a public grant overseen by the French National Research Agency (ANR) as part of the « Investissements d'avenir » program (reference : ANR-10-EQPX-17 – Centre d'accès sécurisé aux données – CASD).

Updated files and series are available on the WID.world website (World Wealth and Income Database): <http://WID.world>. Contacts: Garbinti (Banque de France, Crest): [bertrand.garbinti@ensae.fr](mailto:bertrand.garbinti@ensae.fr); Goupille-Lebret (Paris School of Economics, GATE-LSE): [jonathan.goupille@ens.fr](mailto:jonathan.goupille@ens.fr); Piketty (Paris School of Economics): [piketty@psemail.eu](mailto:piketty@psemail.eu). This paper presents the authors' views and should not be interpreted as reflecting those of their institutions.

[Appendix A. National income and wealth accounts](#)

[Appendix B. Benchmark unified DINA series \(1900-2014\)](#)

[Appendix C. Detailed DINA series using tax micro-files \(1970-2014\)](#)

[Appendix D. Fiscal income series using tax tabulations \(1900-2014\)](#)

This Data Appendix has two main purposes: to provide all relevant details on the data sources and methods we use in this research, and to provide complete data series on income inequality dynamics.

The Appendix is organized as follows. In Appendix A, we present long-term series on income and wealth derived from national accounts. All data files, computer codes, additional series and robustness checks regarding the different methods used in our benchmark income series are presented in Appendix B to D. Appendix B presents complete tables and figures of our benchmark unified series of DINA income distributions (as well as pre-tax labor and pre-tax capital income) for the 1900-2014 period. These series are obtained by combining different data sources and methods over time. Appendix C presents the detailed DINA series using income tax micro-files broken down by age, gender and income categories. In this Appendix, we describe both the income tax returns micro-files and the methodology used to derive the DINA series for the 1970-2014 period. Tables and figures related to different concepts of incomes from 1970 onwards, as well as income composition and income concentration are also detailed. Appendix D present detailed fiscal series based on income tax tabulations from 1900 onwards.

This Appendix is supported by several series of Excel and PDF files as well as computer codes that contain and present our complete income inequality series. The directory GGP2017DINA.zip is organized as follows. For each section of the Appendix, there is a folder called GGP2017DINAAppendixX (with X=A,...,D). Each of these folders contains all the relevant materials (Excel files, computer codes, etc.) as well as a ReadMe file presenting these elements. The Excel files are called GGP2017DINAAppendixX.xlsx and contain all tables and figures relatives to the section and excluded from the main text for the sake of conciseness. These Excel files can be supplemented by a DataFiles folder including all computer codes and raw data used to produce the income inequality series.

## **Appendix A. National income and wealth accounts**

The Appendix A includes tables and figures on the long-term structure of aggregate income and wealth derived from the French national accounts.

Each year, the French national statistical institute (Insee, Institut national de la statistique et des études économiques) publishes retropolated national accounts in a comprehensive, consistent and homogenous manner. For this paper, we have used the most recent edition of the national accounts, i.e the 2015 edition, which follows the 2008 SNA and have 2010 base year. The French national accounts provide a full description of income accounts by sector starting in 1949. Wealth accounts by sector are also constructed starting in end 1969.<sup>1</sup> For the earlier periods, we use the historical series provided by Piketty and Zucman (2014).<sup>2</sup>

Our wealth and income series are updated and refined versions of Piketty (2010, 2011) and Piketty-Zucman (2014). Unless otherwise noted, our concept and methodology are similar to Piketty-Zucman (2014). Our series cover mainland territory and overseas departments, rather than only the mainland territory.<sup>3</sup> We compute real values using the GDP deflator. All our wealth data points are mid-year estimates rather than beginning-of-year estimates.

Compared to the works previously quoted, the following changes have been made and should be pointed out. First, our series of income and wealth have been updated using the most recent version of the national accounts, i.e the 2015 edition using SNA 2008. Second, we extend the national account series by providing a decomposition of pre-tax factor income and pre-tax national income by income categories starting in 1896 (Table A8 to A13) and a decomposition of real yield,

---

<sup>1</sup> Wealth accounts released by INSEE include both non-financial and financial balance sheets. Financial accounts are computed by the Bank of France at a very detailed level and can be downloaded either on INSEE or Bank of France website. The 2015 edition of wealth accounts is available online and starts in end 1978. 1969-1978 data are not available online. They can be found in "25 ans de Comptes de patrimoines (1969-1993)", Insee Resultats, no.348, December 1994, 129p.

<sup>2</sup> See the Data Appendix of Piketty (2010, 2011) and Piketty and Zucman (2014) for a complete description of the historical series of aggregate income and wealth as well as the references used to construct this homogeneous series (Bourguignon and Levy-Leboyer, 1985 and Villa, 1994)

<sup>3</sup> Consistent with Piketty-Zucman (2014), our population series excludes so-called overseas territories (Nouvelle-Caledonie, Polynesie, Wallis-et-Futuna, etc.) and Monaco.

capital gains and investment rate by assets starting in 1970 (Table A23 to A27). Finally, we focus primarily on personal wealth instead of private wealth, i.e. wealth of the household sector excluding wealth from non-profit institutions serving households (NPSIH). Indeed, the purpose of this paper is the study of wealth inequality among individuals. As NPISH represents less than 1 percent of personal wealth, the exclusion or the inclusion of NPISH wealth does not affect significantly our wealth series.

Table A0 presents various long-term series of aggregate national income and personal wealth from 1700 to 2014. The series are depicted either in current or in constant 2014 euros using the GDP deflator index. We also provide per capita or per adult series<sup>4</sup>.

### Section A.1. National income accounts

Tables A1 to A13 provide an overview of the national income circuit from the production to the distribution of primary income from 1896 to 2014. They show also the transition from national income to pre-tax factor income and pre-tax national income.

Table A1 provides the decomposition of the long-term series of national income from 1896 to 2014. National income ( $Y_t$ ) is obtained by subtracting the capital depreciation<sup>5</sup> ( $KDt$ ) and adding the net foreign factor income ( $FY_t$ ) to the gross domestic product ( $GDP_t$ ) such as :  $Y_t = GDP_t - KD_t + FY_t$ .

#### National income circuit from the production to the distribution of factor income

By definition, the national income is equal to either the sum of the net value-added of all institutional sectors or the sum of primary income received by each sector.

<sup>4</sup> Per adult series are obtained by dividing the aggregate national income or personal wealth by the adult population, i.e. individuals over aged 20.

<sup>5</sup> The income accounts computed by INSEE do not include consumption of fixed capital (CFC) by sector prior 1978. Indeed, CFC comes from wealth accounts that starts in end 1978. Consequently, from 1896 to 1977, only gross incomes are reported into the French National Accounts. To deal with this issue, we compute for each sector the ratio CFC/gross value added in 1978 and we assume it constant before 1978.

In the national accounts, six institutional sectors are defined: non-financial corporations, financial corporations, general government, non-profit institutions serving households (hereafter NPSIH), households and rest of the world sectors.

In contrast with national accounts, we define six alternative sectors (hereafter production sectors): housing sector, non-corporate sector, corporate sector, government and NPSIH sector, foreign sector and production taxes sector (see details below).

Tables A2 and A3 present the decomposition of national income between these six alternative sectors. First, we regroup all production taxes net of subsidies received by the government (D2-D3) and the current surplus of the government (B2n) into the production taxes sector (Col 6). The net value-added of the remaining sectors is therefore net of production taxes. Second, the household sector is split into housing sector and non-corporate sector. Housing sector (Col 1) corresponds to the net operating surplus of households (B2n). This surplus corresponds to all net rents from housing assets owned by the households (imputed and real ones) net of production taxes and capital depreciation. Non-corporate sector (Col 2) corresponds to mixed-income from self-employed (B3n)<sup>6</sup> and non-corporate workers (compensations of employees paid by households). Third, we regroup non-financial and financial corporations into a unique corporate sector. The net value-added of corporate sector (net of production taxes) is reported on Col 3. It corresponds to the sum of the net operating surplus of non-financial and financial corporations (B2n) and wages and social contributions paid by the corporate sector (D1). Fourth, government and NPISH sector reported in Col 4 corresponds to compensations of employees paid by the government and the NPSIH sectors (D1). Fifth, we regroup into the foreign sector reported in Col 5 the sum of net foreign labor income and net foreign capital income received from the rest of the world.

Tables A4 to A9 show the transition from the production to the distribution of primary income for each sector.

---

<sup>6</sup> Mixed-income is the economic surplus generated from unincorporated companies. It reflects both labor income (remuneration of the owner of the company) and capital income (returns from the invested capital).

In Table A4, the net-value added of the corporate sector is equal to the compensations of employees and the net operating surplus, which is split into the distributed profit, the corporate income taxes, the retained earnings and the other transfers.

In Table A5, the net-value added of the government and NPSIH is almost entirely equal to the compensation of employees. Indeed, the production value of the public administrations and NPISH are fixed, by convention, to their production cost. The net primary surplus of the government is then equal to the production taxes and the net capital received. Since 1984, the net capital income received by the government is slightly negative. It means that the French government pays more debt interests than it receives capital income from its investments in the economy.

In Table A6, the net-value added of the housing sector is equal to the sum of the imputed rents plus the real rents minus the mortgage interests paid by households. Note that the national accounts do not provide a decomposition of rental income into real rents and imputed rents, i.e. the rents that home owners would have received if they have rented their dwellings. Fortunately, the French ministry of housing computes detailed housing accounts breaking down total rents into real and imputed ones for each year since 1984.<sup>7</sup> For each year since 1984, we then compute the ratio (imputed rents)/(total rents paid to household) that we applied to our series of rental income coming from the national accounts in order to estimate imputed rents. As the ratio is remarkably stable around 75 to 77% for the whole 1984-2014 period, we use the ratio observed in 1984 for the 1949-1983 period.

In Table A7, the net-value added of the non-corporate sector is equal to the sum of the compensations of employees (paid either by households or by non-corporate businesses) and mixed income self-employed. This mixed income is then split into a labor and a capital component using two alternative methods. In the first method, we assume that the non-corporate business sector has the same factor shares (labor vs capital) than the corporate sector. In the second method that we favor, we assume

---

<sup>7</sup> These accounts are available on <http://www.statistiques.developpement-durable.gouv.fr/publications/p/references/compte-logement-2014-premiers-resultats-2015.html>



that labor income represents 70% of mixed income. As it turns out, both methods give estimates that are not significantly different.

### From national income to pre-tax factor and pre-tax national income

Table A8 presents the net primary surplus of households (personal income) by type of income. This surplus is equal to all incomes (net of production taxes and capital depreciation) received by households from other production sectors. It is split into a labor component (compensation of employees and labor component of mixed income) and a capital component (capital component of mixed income, imputed and real rents net of mortgages, interests from debt assets and saving accounts, life insurance income, and dividends<sup>8</sup>).

Table A9 breaks down the national income by total primary surplus of each sector. The national income is equal to the personal income of the households (labor and capital incomes described in Table A8), the net primary surplus of corporate sector (corporate taxed and non-distributed profits) and the primary surplus of the government and NPSIH (total product taxes received by the government, net operating surplus and net capital income).

Table A10 shows which components need to be added to go from the primary surplus of households to pre-tax factor income. Pre-tax factor income is equal to the sum of all pre-tax personal income flows accruing directly or indirectly to the individual owners of the production factors, labor and capital, before taking into

---

<sup>8</sup> Note that the SNA 2008 has introduced a change in the classification of property income flows, which in some cases might have significant consequences for the definition of the different asset-level rates of return. The classification of financial assets was virtually unchanged in the new system, so that we can define our three main categories of financial assets in the same manner in the two systems: deposits, currency, bonds and loans (sum of AF1, AF2, AF3, AF4, AF7, AF8); equity and mutual funds (AF5); life insurance and pension funds (AF6). However the classification of property income flows was changed in a significant way: in the SNA 1993, D44 corresponded to income from life insurance and pension funds. In contrast, D44 corresponds now to "Investment income disbursements" in the SNA 2008 and includes both income from life insurance and pension funds (D441+D442) and "Investment income attributable to collective investment funds share holders" (D443), which was previously included into D42 (together with dividends and other property income flows going to AF5-type financial assets). In order to correct for this, we had access to detailed national accounts and reattribute D443 to the flow of property income going to equity and mutual funds (D42). The sum of D443 and D42 is reported on Table A8, col 17 as dividends such that the flow of dividends corresponds exactly to the asset category "Equities and investment fund shares" reported on Table A20.

account the operation of the tax transfer system and before taking into account the operation of the pension system. In order to construct series of pre-tax factor income consistent with national income, one need to add to the primary surplus of households, the primary surplus of corporations (undistributed profits and corporate income taxes), the net surplus and capital income received by the government and NPISH as well as production taxes.

Table A11 reports pre-tax factor income by income categories once we have added all the components. We apply the following tax incidence assumptions. First, we assume that all production taxes (col 15 "Taxes on production less property taxes") fall proportionally on the different income categories, except for the property taxes (Col 16 "Property taxes") that is fully imputed to rental income. Second, the share of undistributed profits and corporate income taxes accruing directly to households<sup>9</sup> is attributed proportionally to financial incomes excluding interest from saving accounts. Finally, the share of undistributed profits and corporate income taxes accruing to the government as well as the net primary surplus and the net capital income of government and NPISH fall proportionally on the different income categories.

Table A12 shows how to go from pre-tax factor income to pre-tax national income. Pre-tax national income (or more simply pre-tax income) is our benchmark concept to study the distribution of income. Pre-tax national income is equal to the sum of all income flows going to labor and capital, after taking into account the operation of the pension system, but before taking into account other taxes and transfers. That is, we deduct pension contributions (as well as other social contributions, as defined by SNA 2008 national accounts guidelines) from incomes, and add pension benefits (as well as other social benefits, as defined by SNA 2008). When the social contributions do not match the social benefits, we impute the deficit or the surplus proportionally to the whole income.<sup>10</sup> By definition, aggregate pre-tax national income and pre-tax factor income are both equal to national income. However, they differ in terms of

---

<sup>9</sup> Undistributed profit and corporate income taxes are attributed to both the government and the households depending on the relative share of equity owned by each of them. The share of equity owned by the government is equal to 25%-35% over the 1986-2014 period (50% to 60% over 1978-1984 period).

<sup>10</sup> Typically, the social contributions are more important than the social insurance benefits since WWII as they also finance social assistance.

income distribution as the elderly have no labor income in pre-tax factor income while they get pension benefits in pre-tax national income.

Table A13 reports the resulting pre-tax national income broken down by income categories. In contrast with Table A11 (pre-tax factor income), compensations of employees and labor component of mixed income are reported net of social contributions and a new income category corresponding to pensions and unemployment benefits is added.

### Section A.2. Personal wealth accounts

Table A20 presents the evolution of personal wealth, i.e. net wealth owned by households, broken down by asset class over the 1970-2014 period. The net personal wealth is defined as the sum of non-financial assets and financial assets, net of financial liabilities (debt), held by the household sector. We break down non-financial assets into housing assets and business assets. We include in housing assets the value of the building and the value of the land underlying the building. We include in business assets all non-financial assets held by households other than housing assets.<sup>11</sup> We break down financial assets into four categories: deposits (including currency and saving accounts)<sup>12</sup>; bonds (including loans); equities (including investment funds shares); life insurance (including pension funds). We therefore have seven asset categories (housing assets, business assets, four financial asset categories, and debt), or actually eight categories when we break down housing into owner-occupied and tenant-occupied housing.<sup>13</sup>

In Table A21, we divide all the assets by national income in order to get wealth-income and assets-income ratio. Finally, Table A22 reports the decomposition of personal wealth by asset in percentage of total wealth.

### Rates of returns, rates of capital gains and savings-induced wealth growth rate

---

<sup>11</sup> In practice, these are mostly the business assets held by self-employed individuals (but this also includes other small residual assets).

<sup>12</sup> Saving accounts correspond to AF2-AF292-AF293 in the financial accounts of the national accounts.

<sup>13</sup> The national accounts do not provide a distinction between owner-occupied and tenant-occupied housing. We use the housing accounts compiled by the French ministry of housing to compute it.

Tables A24a and A24b report the annual flow returns on personal wealth by asset class. These returns are obtained by dividing each pre-tax capital income<sup>14</sup> (reported on Table A11) by the corresponding asset (reported on Table A21). Two estimations have been made regarding the treatment of retained earnings and corporate income taxes. In Table A24a, retained earnings and corporate taxes accruing to household are added to flows from bonds, equities, and life insurance and pension funds. In Table A24b, retained earnings and corporate taxes accruing to households are entirely added to equity flow. We tend to favor Table A24a over Table A24b as our benchmark scenario. Indeed, retained earnings and corporate taxes should not fall only on corporate equity but also on other forms of financial assets.<sup>15</sup> Finally, Table A24c reports the annual flow returns on personal wealth by asset class when considering after-tax capital income instead of pre-tax capital income.

Table A26 reports the savings-induced wealth growth rate and real growth rate by asset class. We can formerly write the accumulation process of an asset A from year t to year t+1 as:

$$A_{t+1} = (1 + p_t)(1 + g_{wt})A_t \quad (1)$$

$$A_{t+1} = (1 + p_t)(1 + q_t)(1 + g_{wst})A_t \quad (2)$$

Where  $q_t$  is the asset-specific rate of real capital gains at time t,  $p_t$  is the inflation rate using the GDP deflator index and  $g_{wst}$  is the savings-induced wealth growth rate.

The real growth rate is simply equal to the increase or decrease of the given asset once corrected for inflation ( $g_{wt} = \frac{A_{t+1}}{(1+p_t)A_t} - 1$ ). The savings-induced wealth growth rate  $g_{wst}$ , is the yearly variation of the aggregate asset resulting from new investments and capital depreciation. These savings-induced wealth growth rate can be easily computed for each asset using the French national wealth accounts.<sup>16</sup> Indeed, these accounts report the yearly variation of each asset resulting from both price effect, i.e.  $(1 + p_t)(1 + q_t)$ , and volume effect  $(1 + g_{wst})$ . Note that we make

---

<sup>14</sup> The pre-tax capital income is gross of production taxes. It also includes the primary surplus of corporate sector (retained earnings and corporate income taxes). See Table A11 and the description of this table in the Appendix for more details.

<sup>15</sup> Harberger (1962)'s seminar analysis showed that asset owners arbitrage taking into account differences in the net-of-tax returns.

<sup>16</sup> Using the « Compte de variations du patrimoine ».

only one adjustment to these series. In the national wealth accounts, retained earnings and net capital transfers are considered as capital gains because they do not correspond to investments made by households. In fact, they represent corporate savings, i.e. indirect investments made by households, and not capital gains. So, we add this share of corporate savings accruing to households to the investment/savings made by households.

Tables A25a, A25b and A25c report the real rates of capital gains on personal wealth by asset class. These rates of capital gains are obtained for each asset using Equation 2 (above) such that:

$$1 + g_{wt} = (1 + q_t)(1 + g_{wst})$$

Three estimations have been made regarding the treatment of corporate savings. In Table A25a, which is our benchmark scenario, corporate savings are added to investments/savings made by households in all financial assets except deposits. In Table A25b, they are added entirely to household investments in equities. In Table A25c, they are excluded and therefore considered as capital gains.

The three Tables A23 report the total returns on personal wealth by asset class for the three scenarios a, b and c. They are simply the sum of the flow returns (reported in Tables A24) and of the real rates of capital gains (reported in Tables A25).

Table A27 reports the decomposition of personal wealth accumulation between volume and price effects over the 1896-2014 period. More precisely, this table shows the evolution of the real growth rate of personal wealth that can be split into a volume effect (savings-induced wealth growth rate) and a price effect (real rate of capital gain), along with the flow rate of return (excluding capital gains) and the total rate of return (including capital gains).

### Section A.3. Private, Public and National wealth

Table A28a presents the evolution of private wealth by asset class over the 1970-2014 period. The private wealth is equal to the sum of the personal wealth, i.e. net

wealth owned by households, and the wealth of the NPSIH. Table A28b reports the same decomposition but divided by national income.

Table A29 shows the evolution of public wealth, corporate wealth and net foreign wealth by asset class over the 1970-2014 period. The net foreign wealth corresponds to the foreign assets owned by French residents minus the French assets owned by foreigners. The corporate wealth is computed either at its book value, i.e. the sum of non-financial assets plus financial assets minus financial (non-equity) liabilities, or at its market value, i.e. the financial equity liabilities. The Tobin's Q is reported on col 12 as the ratio of the market value over the book value. Note that in France, Tobin's Q has always been lower than 1 and the difference between the two concepts has been pretty stable, around 20% of national wealth. Table A29b shows the same decomposition but in percentage of national income. Table A30a and A30b report the evolution of national wealth. National wealth can be computed either at its book value or at its market value.

## **Appendix B. Benchmark unified DINA series (1900-2014)**

In this Appendix, we present complete tables and additional figures of our benchmark unified income distributions series for the 1900-2014 period. We use as a benchmark the series of pre-tax national income<sup>17</sup> among equal-split adults.<sup>18</sup> Let us first insist on the fact that there is no unique way to represent income inequality. Different concepts of income (fiscal income excluding or including capital gain, pre-tax factor and pre-tax national incomes, etc.) and population units (tax units, equal-split adults, individualistic adults<sup>19</sup>) can be used to highlight complementary aspects of income inequality.

We discuss in this Appendix (as well as in Appendixes C and D) the rationales behind the choice of our benchmark concepts, i.e. pre-tax national income among equal-split adults, and when alternative concepts can be more suitable. We then present the methodology used to construct homogenous series of pre-tax national income over the 1900-2014 period. Finally, we compare our benchmark series with alternative series using different concepts of income and population units.

### **Section B.1. Concepts of income and population units**

#### **Fiscal income vs Pre-tax national and Pre-tax factor income**

Previous work on long run evolution of income inequality has focused on fiscal income among tax units (Piketty, 2001, 2003). As described in the main paper, we define and use two alternative concepts of income that are 100% consistent with national income (pre-tax national income and pre-tax factor income) and therefore more suitable to study income inequality.

---

<sup>17</sup> As precised below, pre-tax national income is equal to the sum of all income flows going to labor and capital, after taking into account the operation of the pension system, but before taking into account other taxes and transfers.

<sup>18</sup> i.e. the income of married couples is divided into two equal parts.

<sup>19</sup> i.e. labor income is allocated to each individual income earner within the couple. Note that capital income of married couples is always divided into two (because we do not have other information).

Pre-tax national income is equal to the sum of all income flows going to labor and capital, after taking into account the operation of the pension system, but before taking into account other taxes and transfers.

Pre-tax factor income is equal to the sum of all income flows going to labor and capital, before taking into account the operation of the pension system, and before taking into account other taxes and transfers. That is, we do not deduct pension contributions (or other social contributions).

The main limitation of fiscal income that was widely used in previous work is that it can dramatically diverge from national income due to non-taxable components of income and production taxes (see Appendix C for more details on the way we impute these missing components for the 1970-2014 period). We detail below the reasons why discrepancies may appear for both capital and labor incomes.

Fiscal capital income can dramatically diverge from national-accounts capital income<sup>20</sup> since a significant part of capital income is tax-exempt and therefore not reported in income tax data (See Figure FA1 in Appendix A). This is the case of owner-occupied rental income (imputed rent); interest and dividend income going to tax-exempt life insurance assets (life insurance income); other tax-exempt interest income paid to deposits and saving accounts.

Corporate retained earnings and corporate taxes are also not included into fiscal capital income and need to be imputed in order to be consistent with national capital income. Indeed, corporate taxes are paid by corporations before they can distribute capital income flows, and are therefore excluded from fiscal income. Corporate retained earnings, while they constitute a source of capital income to individuals, are not directly distributed to individuals and therefore not included into fiscal capital income.

It is worth stressing that all these components have increased significantly in recent decades. In particular, life insurance assets did not play an important role until the

---

<sup>20</sup> pre-tax capital income equals factor capital income as these concepts differ only because of the treatment of pension contributions.



1970s, but gradually became a central component of household financial portfolios since the 1980s-1990s. Imputed rent has also become more and more important over time because of the rise of homeownership. Finally, corporate retained earnings and corporate taxes were relatively small until the mid-20<sup>th</sup> century and also increased significantly in recent decades.

Fiscal labor income is different from both factor labor income and pre-tax labor income. It differs from factor labor income for two main reasons. First, the social security contributions (employer and employee social contributions) are not taxable and therefore not reported into the income tax returns. Second, fiscal labor income includes replacement incomes (pension and unemployment benefits), although these incomes are not factor incomes and should therefore be equal to 0 in this framework.

Fiscal labor income differs from pre-tax labor income due to non-taxable compensation items such as health benefits and a number of other in-kind benefits. Finally, fiscal income also differs from pre-tax factor and pre-tax national income because of production taxes (in the SNA 2008 sense). These production taxes include a number of indirect taxes, such as value added taxes, which are paid by corporations before they can distribute labor and capital income flows, and are therefore excluded from fiscal income.

Taking into account all these non-taxable components is crucial to measure and compare adequately income inequality not only over time but also across countries. This is why we favor pre-tax national income and pre-tax factor income over fiscal income.

Our benchmark income concept is pre-tax national income. We prefer it over pre-tax factor income since an important limitation of the concept of factor income is that, by definition, retired and unemployed individuals have no factor labor income. Therefore, inequality of pre-tax factor income tends to rise mechanically with the fraction of retired individuals in the population, which biases comparisons both over time and across countries. Pre-tax factor income must then be only used to study income inequality among working-age populations (e.g. individuals aged 25-59 year-old) rather than among overall population. We also argue that looking at the distribution of

factor labor income among the working-age population can be insightful to better measure the distribution of labor costs paid by employers.

### Tax units vs equal-split adults vs individualistic adults

We now present the three different concepts of population units that can be used. The distribution of income can be shown among tax units (i.e. married couples and singles), among equal-split adults (i.e. the income of married couples is divided into two equal parts) or among individualistic adults (i.e. labor income is allocated to each individual income earner within the couple<sup>21</sup>).

Representing income inequality among tax units can be problematic since tax units mixed both couples and singles, creating artificial inequality between these two different types of observation unit. This could be a concern when comparing income inequality over time and across countries for various reasons. First, the share of singles has continuously increased in France since the early 90s, increasing artificially the level of income inequality across tax units (see Figures FD13 to FD14, Appendix D for a comparison). Second, the incentives for filling jointly or separately an income tax return may differ across countries, creating comparability issues.

The concept of equal-split adult as unit of observation allows us to overcome these limitations and is therefore our benchmark population unit. It constitutes a simple way to correct for differential size of tax units, by dividing the income of married couples into two equal parts. Of course, there are other possible ways to split income among couples, by taking into account marital specialization, distribution of home and care duties, economies of scale, etc. However, these issues are far beyond the scope of this work. We do not try here to take into account these elements and simply correct for differential size of tax units.

Nevertheless, we argue that the concept of individualistic adult as unit of observation is more suitable when studying gender inequality. In this particular case, individualized series of labor income (i.e. when each spouse is assigned his or her

---

<sup>21</sup> Note that capital income of married couples is always divided into two (because we do not have other information).

own labor income) can highlight changes in woman (and man) labor force participation and possible glass ceiling.

To sum up, we use as a benchmark our series of pre-tax national income among equal-split adults to measure income inequality. However, we stress that alternative concepts of income or unit of observation can be more suitable in some particular cases. Factor labor income among working-age individualistic adults can be useful to better measure the distribution of labor costs paid by employers. Pre-tax national income among individualistic adults can be more suitable when studying gender inequality.

### Section B.2. From fiscal income to pre-tax national income over the 1900-2014 period

We now present how our series of pre-tax national income among equal-split adults are constructed over the 1900-2014 period.

First, we combine income tax micro-files, household surveys (housing and wealth surveys) and national accounts to provide series of pre-tax national income by population unit (tax units, individualistic adults and equal-split adults) and by income categories (labor and capital) for the 1970-2014 period. The data and the methodology used are explained in details in Appendix C.

Second, we estimate long-term series of fiscal income by population unit (tax units and equal-split adults) for the 1900-2014 period using income tax tabulations. The data and the methodology used are explained in details in Appendix D.

Finally, we estimate long-term series of pre-tax national income by population unit (tax units and equal-split adults) and by income categories (labor and capital) over the 1900-2014 period using the long-term series of fiscal income over the 1900-2014 period from Appendix D and the series of pre-tax national income over the 1970-2014 period from Appendix B. Below, we go into further detail in the presentation of the series of pre-tax national income as well as in the methodology used.

### Unified series of pre-tax national income over 1900-2014

Tables B1 and B2 present descriptive statistics of our benchmark series of pre-tax national income among equal-split individuals from 1900 onwards. Figures B1 to B4 show income shares in France for the same period and for the same benchmark concept of income, from the bottom 50 % to the top 0.1% of the income distribution.

Tables B3 and B4 show descriptive statistics of pre-tax national income among tax units rather than among equal-split adults. We compare the level of income concentration derived from series of pre-tax national income among either tax units or equal-split adults as unit of observation in Figures B8 and B9. Moving from tax-unit series to equal-split series decreases slightly the level of income concentration, given the rise of the fraction of singles since the early 1990s.

In order to estimate the distribution of pre-tax national income from the distribution of fiscal income, we proceed in two steps.

First, for the 1970-2014 period, we take advantage of the fact that we can compute both series of fiscal income (based on income tax micro-files) and series of pre-tax national income (combining income tax micro-files with sophisticated imputation procedures based on national accounts and household surveys). The data and the methodology used to estimate series of pre-tax national income over the 1970-2014 period are explained in details in Appendix C. We naturally use these corrections in order to construct our benchmark series of pre-tax national income. That is, we compute the national-income/fiscal-income ratios by year and percentile using the micro-files series, and we apply these ratios to the fiscal-income tax-tabulations series.<sup>22</sup> The systematic comparison for the 1970-2014 period between the distribution of fiscal income (from bottom to top percentiles) estimated via the micro-files and via the income tax tabulations shows that the difference between the two

---

<sup>22</sup> The national-income/fiscal-income ratios for the period 1970-2014 are computed by using Tables C4, C6, C13 and C15 of Appendix C. The Stata code used to estimate series of pre-tax national income from series of fiscal income among equal-split adults and tax units is reported in the folder StataFiles from Appendix B (do\_gperc\_income).

series is negligible (Figures C1 to C3 from Appendix C) and gives us confidence in the validity of our methodology.

Second, for the 1900-1970 period, our correction procedure is more rudimentary. We start from the presumption that the induced corrections on percentile shares tends to rise over time. This assumption is motivated by the fact that at the beginning of the period, tax rates are relatively small, so that incentives for tax optimization are limited, and legal tax exemption regimes are rare. This is confirmed by the detailed breakdowns by labor and capital incomes (which unfortunately are only available for a number of isolated years – rather than on an annual basis – prior to 1945; see Piketty 2001 and on-line appendix files). Then, we assume that our national-income/fiscal-income ratios rise linearly from 1 in 1900, i.e. inequality in pre-tax national income is equal to inequality in fiscal income, to the ratios observed in 1970.

Let us insist on the fact that this is clearly an approximation, but the impact on our long run patterns is very limited (and in any case would tend to reinforce our main findings). Indeed, Figures B5 to B7 show that the difference between the series of pre-tax national income and fiscal income is negligible until the mid-1980s. As expected, we observe a growing gap between the two series at the top of the distribution since the mid-1980s. As already discussed above, it reflects the growing importance of missing capital income and retained earnings in fiscal income that are mainly concentrated at the top of the distribution.

Finally, note that we do not attempt to provide either pre-tax factor income series or fully individualized series prior to 1970. Indeed, tax tabulations do not include any information on within-couple distribution of income, so one would need to find other data sources in order to do this. More generally, we stress that our long-run series should be viewed as exploratory and incomplete, and we hope that they will be further developed and refined in future research.

#### Labor and capital incomes series

Tables TB9 and TB11 describe respectively the series of pre-tax labor income and pre-tax capital income among equal-split adults. These long-term series are obtained

by combining our series of pre-tax labor and capital income over the 1970-2014 period (from Appendix C) with the long-term series of fiscal labor and capital income.

For the 1970-2014 period, we directly used the series of pre-tax labor and capital income reported in Tables C22-C23 (for labor income) and C24-C25 (for capital income) from Appendix C. These series are obtained by combining income tax micro-files with sophisticated imputation procedures based on national accounts and household surveys. The data and the methodology used to estimate these series over the 1970-2014 period are explained in details in Appendix C.

For the 1900-1970 period, we use the fact that income tax tabulations include detailed breakdowns by income categories (wages, self-employment income, dividend, interest, etc.), to estimate separately the distribution of fiscal labor income and fiscal capital income. As noted before, one important limitation of the detailed tabulations by income categories is that, prior to 1945, they only cover a limited number of years (namely, 1917, 1920, 1932, 1934, 1936 and 1937); they then become annual in 1945. Fortunately there are separate annual tabulations for wages over the 1919-1938 period, and quasi-annual inheritance tabulations over the 1902-1964 period.

For labor income (Table TB9), regarding the 1919-1969 period, we start from long-term top wage series derived from income tax tabulations and provided by Piketty (2001). We then adjust them to obtain pre-tax labor income series. The simplest way to do this is to compute the ratio between the top wage shares and our benchmark pre-tax labor income series in 1970 and to assume the ratio constant for the period 1900-1969. Indeed, the ratios reported in Table B9 are not very important (1.1 for P90-100 and 1.01 for P99-100) and very stable over time (the ratios observed in 1970 is similar to the average ratios computed for the 1970-1985 period and very close to those computed over the 1970-2014 period). For the 1900-1918 period, we use the evolution observed for the pre-tax national income top shares to interpolate linearly the previous shares.

For capital income, we use our top wealth share series (presented in our companion paper<sup>23</sup>) to linearly interpolate the top pre-tax capital income shares from 1900 to 1970, using the difference between top capital income shares and top wealth shares.

Finally, Figures B10 to B13 compare our long-term series of wealth, pre-tax national income, pre-tax capital income and pre-tax labor income over the 1900-2014 period. Total income concentration appears much more pronounced than labor income one, which points out the major role of capital income in total income inequality. It also shows that the long run decline in income inequality is due to the fall of top capital incomes following the 1914-1945 capital shocks. As expected, capital income and wealth are much more concentrated than labor income.

---

<sup>23</sup> Top wealth share series are depicted in Table A1, Appendix A of our companion paper (Garbinti, Goupille-Lebret, Piketty, 2016).

## **Appendix C. Detailed DINA series using income tax micro-files (1970-2014)**

The Appendix C relates to the estimation of detailed income distribution series broken down by percentile, age, gender and income categories over the 1970-2014 period. The series are obtained by combining three main types of data: national accounts; fiscal data (micro-files of income tax returns); and household surveys (wealth and housing surveys). In this Appendix, we first present briefly the raw data sources used (income tax micro-files and household surveys) as well as the methodology to estimate income series consistent with national income. We then present the supplemental tables and figures relative to our detailed series of income.

### Section C.1. Data sources and methods

#### Fiscal income and micro-files of income tax returns

As described in the main paper, the estimation of the income distributions for the 1970-2014 period is based on micro-files of income tax returns. These micro-files have been produced by the French Finance Ministry since 1970 and fall into two categories: “Enquête Revenus Fiscaux” (Tax Income surveys, hereafter: ERF surveys) and “Echantillons Légers et Lourds” (hereafter: samples of income tax returns).

We use the first series of ERF surveys produced jointly by Insee<sup>24</sup> and the tax administration every 5 years from 1970 to 1990.<sup>25</sup> The surveys describe the socio-demographic structure of approximately 40,000 tax units along with all the information reported in their income tax returns (containing different sources of taxable income and income tax).

---

<sup>24</sup> Insee stands for Institut National de la Statistique et des Études Économiques and is the national institute in charge of the production, the analysis and the diffusion of official statistics in France.

<sup>25</sup> The first series of ERF surveys was edited eight times since 1956 (1956, 1962, 1965, 1970, 1975, 1979, 1984 and 1990). The first ERF of 1956, 1962 and 1970 are not available anymore. The Tax Administration was responsible for filling the data related to tax income, while Insee was in charge of the statistical data processing. The updated version of these surveys is now called The Tax and Social Incomes Survey (ERFS). These surveys are annual and match information from Labor Force surveys with income tax returns and social benefits perceived. See description of Tax Income Survey/ERF and Tax and Social Income Survey/ERFS on Insee website.



In addition, we have access to large samples of income tax returns edited each year by the tax administration since 1988. These files include 40,000 tax units from 1988 to 1993 (Echantillon léger) and about 400,000-500,000 tax units per year since 1994 (Echantillon lourd). These micro-files are stratified by taxable income brackets with large oversampling at the top (they are exhaustive at the very top). Since 2010 we also have access to exhaustive micro-files, including all tax units, i.e. about 37 million tax units in 2010-2012.

These micro-files allow us to estimate directly the distribution of fiscal income, i.e. income reported on income tax returns, by income categories. In particular, fiscal labor income can be split into three components (wages; pension and unemployment benefits; and labor component of mixed income, which we assume for simplicity to be equal to 70% of total mixed income<sup>26</sup>) and fiscal capital income into four components (tenant-occupied rental income; dividend; interests from debt assets; and capital component of mixed income, i.e. 30% of total mixed income).

In order to estimate the distribution of national income (pre-tax or factor), we need to combine income tax micro-files with other data sources, namely national accounts and household surveys, and to apply a number of imputation rules that we describe below.

### From fiscal labor income to DINA factor labor income

Factor labor income is equal to the sum of all income flows going to labor before taking into account the operation of the pension system and before taking into account other taxes and transfers. Fiscal labor income differs from factor labor income for three main reasons. First, the social security contributions (employer and employee social contributions) and the deductible part of CSG<sup>27</sup> are not taxable and therefore not reported into the income tax returns. Second, fiscal labor income includes replacement incomes (pension and unemployment benefits), although these

<sup>26</sup> See the description of Table A7 in Appendix A above for a discussion of the split of mixed income between a labor and a capital component.

<sup>27</sup> CSG stands for « Contribution sociale généralisée ». As compared to the regular income tax, CSG is a withholding tax implemented in 1991, with fixed tax rates by income categories (8.2% for capital income, 7.5% for earnings, 6.6% for pension benefits, and 6.2% for unemployment benefits in 2015). Only a part of this tax, called “CSG deductible”, is deductible from the regular income tax base.

incomes are not factor incomes and should therefore be equal to 0 in this framework. Finally, once social security contributions are added and replacement incomes are subtracted from fiscal labor income, one need to add the production taxes accruing to labor income in order to be consistent with national labor income.<sup>28</sup>

Starting from the income tax micro-files and the distribution of fiscal labor income, we apply the following steps in order to estimate series on factor labor income.

First, we compute the social contributions and the deductible CSG at the individual level in order to get the total employee compensation (labor cost). This computation exercise has to be made at the individual level, rather than simply imputing all social contributions in proportion to fiscal labor income. Indeed, social contributions are not strictly proportional to labor income and often involve significant exemptions for low wages or high wages, with important variations over the 1970-2014 period.<sup>29</sup> In particular, we simulate the different types of employee and employer social contributions: contributory contributions for pensions and unemployment insurance, non-contributory contributions for family and health care. We also apply the relevant regressive social contribution schedules that vary by earnings brackets (less than the social security threshold, between one and four thresholds, between four and eight, above eight times the social security threshold) and over time.<sup>30</sup> Finally, we take into account and simulate the different schemes relative to the cuts in employer's social security contributions for low-wage workers implemented in France since 1993.

---

<sup>28</sup> The production taxes include a number of indirect taxes (such as value added taxes) which in effect are paid by corporations before they can distribute labor and capital income flows, and are therefore excluded from fiscal income.

<sup>29</sup> Our computations rely on parameters from « *Barèmes IPP : prélèvements sociaux, Institut des politiques publiques, avril 2015* ».

<sup>30</sup> Normally, the social contribution schedules vary whether the wage earner is an executive, and whether he works in the private or in the public sector. For the sake of simplicity, we apply the social contribution schedule specific to executives in the private sector to every workers.

Second, we add the deductible CSG and all social security contributions to fiscal wages and fiscal mixed income. We then split the mixed income into a labor component (70%) and a capital component (30%). At this stage, total compensation of employees (wages + social security contributions) and total mixed income (capital and labor components) are consistent with the macro aggregate from the national accounts reported in Table A8 of Appendix A.<sup>31</sup>

#### From factor labor income to pre-tax labor income

As mentioned earlier, pre-tax labor income is equal to the sum of all income flows going to labor, after taking into account the operation of the pension system, but before taking into account other taxes and transfers. The key difference between pre-tax factor income and pre-tax national income is the treatment of pensions and unemployment benefits.<sup>32</sup> In pre-tax factor income, pensions and unemployment benefits are set to zero and wages and mixed income are gross of social contributions. In pre-tax national income, the wages and mixed incomes are net of social contributions and the corresponding social benefits, i.e. pensions and unemployment benefits, are added.<sup>33</sup> Note that there is always a discrepancy between social contributions on one hand and pension and unemployment benefits on the other hand (Table A12, col 7).<sup>34</sup> This is due to the fact that social contributions

---

<sup>31</sup> Before any adjustments, the ratio of our estimated total employee compensation over that from the national accounts (reported in Table A8) ranges from 90% to 99% over the 1970-2014 period. We then adjust proportionally the different estimated series (gross wage, employee social contributions, employer social contributions, cuts in employer social contributions, mixed income and social contributions on mixed income) in order to match the corresponding macro aggregate from the national accounts.

<sup>32</sup> See description of Table A12 for more details on the transition from pre-tax factor income to pre-tax national income.

<sup>33</sup> Fiscal pensions and unemployment benefits reported in the income tax returns are net of deductible CSG. As for pre-tax factor income, we estimate and add the deductible part of CSG to fiscal pension and unemployment benefits. We then add the corresponding production taxes (excluding property taxes) and the surplus from the government and NPSIH.

<sup>34</sup> In particular, the discrepancy between social contributions, and pension and unemployment benefits was equal to 7% over the 1975-1997 period and decreases dramatically to 3% after 1998. This decrease is due to the 1998 reform that shifts partially the financing of health benefits from social contributions (based on labor income) to CSG based on all incomes (labor and capital).

also financed – in addition to pension and unemployment benefits – other social benefits such as health or family benefits and a number of other in-kind benefits. In the absence of specific information, we simply impute them in proportion to all incomes (labor and capital).

See Tables A10 to A13 and the description of these tables in Appendix A (above) for a detailed presentation of the different series (pre-tax national and pre-tax factor incomes) at the aggregate level, and the incidence assumptions made. See also the dofile *revtravcn* in Appendix B of our companion paper Garbinti, Goupille-Lebret and Pikety (2016) for more details on the different steps implemented to compute factor labor income and pre-tax labor income from fiscal labor income at the individual level.

#### From fiscal capital income to national capital income

The reconciliation of capital income (reported in the micro-files) with its counterpart in the national accounts raises more complicated issues than for labor income. Indeed, a significant part of capital income is tax-exempt and therefore not reported in income tax data. This is the case of owner-occupied rental income (imputed rent); interest and dividend income going to tax-exempt life insurance assets (life insurance income); other tax-exempt interest income paid to deposits and saving accounts (See Figure FA1 in Appendix A).<sup>35</sup>

The reconciliation exercise is fully described in the Data Appendix of our companion paper (Appendix B, Garbinti, Goupille-Lebret and Piketty 2016) and in the dofile *revcapwealthcn*. This exercise is conducted in two steps.

The first step is relative to the following tax-exempt incomes: owner-occupied rental income, life insurance income and interest from saving accounts. We first impute the corresponding assets, i.e. life insurance assets, owner-occupied housing assets and deposits (including currency and saving accounts) using available household

---

<sup>35</sup> Corporate retained earnings, corporate taxes and production taxes are also not included into fiscal capital income and need to be imputed in order to be consistent with national capital income.

surveys.<sup>36</sup> We then attribute the corresponding asset income flows on the basis of average rates of return observed in national accounts for this asset class (Table A24a).

More specifically, the imputation procedure is the following.<sup>37</sup> First, in the household surveys, we define groups according to three dimensions: age, financial income, and labor and replacement income (approximately 200 groups for the imputation of owner-occupied housing asset). Second, for each group and each kind of asset to be imputed (owner-occupied housing, deposits, and life insurance), we both compute an extensive margin (the proportion of individuals holding the asset considered) and an intensive margin (share of the total asset owned by the group). Third, in our income tax micro files, we define groups according to the same dimensions (age, financial and labor incomes). Then, within each of these groups, we randomly draw tax units that own the asset accordingly to the corresponding extensive margin (i.e. computed for the asset and the group considered). The intensive margin is then used to impute the asset amount within the asset holders of this group.

The second step is relative to capital income components reported in the income tax micro-files, i.e. tenant-occupied rental income; dividend; interests from debt assets; and capital component of mixed income (i.e. 30% of total mixed income). The reconciliation consists in adjusting proportionally each of these capital incomes in order to match its counterpart in the national accounts (reported in Table A8).<sup>38</sup>

### Production taxes and corporate earnings

The final step of the reconciliation consists in adding to our labor and capital income series the missing components – income indirectly perceived by individuals (retained earnings) and taxes (production taxes and corporate income taxes) – in order to

---

<sup>36</sup> We use the housing surveys conducted by INSEE in 1970, 1973, 1978, 1984, 1988, 1992, 1996, 2001, 2006 and 2013; and the wealth surveys conducted by INSEE in 1986, 1992, 1998, 2004, 2010 and 2014.

<sup>37</sup> See the Appendix B of our companion paper Garbinti, Goupille-Lebret and Piketty (2016) and the dofile revcapwealthcn for a complete and exhaustive description of the imputation method.

<sup>38</sup> That is, we multiply each individual capital income component reported in the micro-files by the corresponding national-income/fiscal-income ratio.

match the macro aggregates of pre-tax factor and pre-tax national incomes reported in Tables A10 to A13.

As discussed in more details in Appendix A, Table A11, we made the following incidence assumptions. First, we assume that all production taxes (col 15 “Taxes on production less property taxes”) fall proportionally on the different income categories, except the property taxes (Col 16 “Property taxes”) that is fully imputed to rental income. Second, the share of undistributed profits and corporate income taxes accruing directly to households<sup>39</sup> is attributed proportionally to financial incomes excluding interests from saving accounts. Finally, the share of undistributed profits and corporate income taxes accruing to the government as well as the net primary surplus and the net capital income of government and NPISH fall proportionally on the different income categories.

### Section C.2. Presentation of the additional tables and figures

Tables C1 to C27 present either summary or detailed statistics for various concepts of income over the 1970-2014 period. The statistics computed are income shares, average income, income thresholds and inverted Pareto coefficient either for broad income groups (summary statistics) or for the 127 generalized percentiles (detailed statistics). The concepts of income are fiscal income, pre-tax national income (all, labor or capital) and pre-tax factor labor income.<sup>40</sup> For each of these income concepts, we break down our series into three population types: tax units, individualistic adults and equal-split adults.<sup>41</sup> In Tables C30 to C37, we break down our pre-tax and factor labor income series by gender.

---

<sup>39</sup> We subtract the fraction of domestic corporate capital (undistributed profits and corporate income taxes) that can be attributed to the government. The remaining part should be attributed to individuals and we thus impute it. See description of Table A10 in Appendix A for more details.

<sup>40</sup> Note that for factor labor income, the series are always restricted to 25-60 years old individuals as retired individuals have factor labor income set to zero by definition. See Section C.1 above for a description of these concepts.

<sup>41</sup> For equal-split adults, the income of married couples is divided into two. For individualistic-adults, labor income is allocated to each individual income earner within the couple. Capital income of married couples is always divided into two (because we do not have other information. See section 3.1 of our core paper for a presentation and the interpretation of these concepts).

Figures C1 to C5 compare the level of income concentration over the 1970-2013 period when considering different income tax data (micro-files vs income tax tabulations) or different income concepts (fiscal income vs pre-tax national income).

Figures C20 to C30 report either income composition by income groups for different years (1970, 1984, 2000 and 2012) or the evolution of income shares by income categories for various income groups (Bottom 50%, Middle 40%, Top 10%, Top 1%,etc.).

## **Appendix D. Fiscal income series using income tax tabulations (1900-2014)**

In this appendix, we present our fiscal income series obtained by using generalized non-parametric Pareto interpolation techniques developed by Blanchet, Fournier and Piketty (2017) to the historical income tax tabulations available in France over the 1900-2014 period.

The Excel File GGPDINAAppendixD includes the main appendix tables and figures on long-term series of income concentration. The file also provides supplemental tables and figures documenting step by step how series are constructed from raw tabulations and allowing comparison between income concepts. An index is included in the file for a complete list of tables and figures. The folder IncomeTaxTabulations includes all references required to compute our income series from income tax tabulations. The folder StataFiles includes Matlab codes and intermediate tables used for our computations.

We present in this section the income tabulation sources used and the several steps of the approach implemented in order to obtain the income distribution series over 1900-2014.

### **Section D.1. Income tax tabulations in France (1900-2014)**

A general income tax was enacted in France in 1914. It took effect for the first time in 1915 (i.e., taxpayers reported their 1915 incomes at the beginning of 1916), and it has applied every year ever since. Most important, the French tax administration has been compiling every year since 1915 (including the World War II era) summary statistics based on the tabulation of all individual income tax returns. The raw materials produced by the tax administration have had the same general form since 1915: the tabulations indicate the number of taxpayers and the amount of their taxable income as a function of a number of income brackets (the number of brackets is usually very large, especially at the top of the distribution). This basic table is available for each single year of the 1915–2014 period.



These tabulations were first used in a systematic manner by Piketty (2001, 2003). In the present paper we update and considerably refine these estimates (see below). We also use estimates of the distribution of income for years 1900 and 1910 that were produced by the French Finance Ministry in the context of the parliamentary debates about the creation of an income tax (using data from various sources, including property taxes and inheritance taxes).

## **Section D.2 Long-run series using historical income tax tabulations (1900-2014)**

For the 1985-2013 period, we apply the generalized, non-parametric Pareto interpolation techniques developed by Blanchet, Fournier and Piketty (2017) to the income tax tabulations in order to produce annual series of fiscal income for the entire distribution and not only to the top decile. These series refine the initial estimates by Piketty (2001, 2003) which did not attempt to go below the 90<sup>th</sup> percentile. Here we are able to analyze the evolution of the full distribution of income.

### **Correction for non-taxable households**

Before 1985, income tax tabulations include only taxable households.<sup>42</sup> The first step consists in imputing the missing non-taxable income accruing to the top 10% of the income distribution, based on corrections from Piketty (2001, 2003). These corrections use the fact that the relevant exemption thresholds vary with the marital status and numbers of children and are fully described in Piketty (2001). The second step consists in estimating the bottom 90% income distribution. We need to make assumption about how to split P0-90 between P0-50, P50-90 and about the evolution of P10 and P50. The simplest assumption is to assume simple proportional split. Computations done to estimate these shares are detailed in Table TD3 in Excel File GGPDINAAppendixD.

---

<sup>42</sup> See Figure FD1 and Table TD1.

This delivers P50 around 70-75% of average throughout postwar period, which is consistent with 1956-1962-1965-1970 estimates from INSEE ERF surveys<sup>43</sup> (see Piketty 2001, p.702-703). Wage inequality series also show clear rise of P90/P10 and P50/P10 over 1950-1968, and strong decline over 1968-1983 (see Piketty 2001, p.670-672), which we reproduce here. In order to match low bottom 50% share in 1970-1975 from DINA micro-files we assume simple linear adjustment over 1970-1980. In order to match higher P0-50 share in 1950 (suggested by wealth inequality estimates), the linear adjustment is taken over the 1950-1970 period. Full data files and computer codes are provided in the files StataFiles and IncomeTaxTabulations.

### Taxable vs fiscal income

One important limitation of these annual income tax tabulations is that they are based upon a concept of “taxable income” (i.e. fiscal income minus a number of specific deductions instituted by the tax law, such as the 10% lump-sum deduction for professional expenses of wage earners – up to ceiling – ; the additional 20% deduction for wage income – also up to ceiling –; the deduction of previous-year income tax that was allowed before 1945; etc.) rather than the concept of “fiscal income” that we are interested in (i.e. income reported on fiscal declarations, before any further deduction). It is critical to correct for these deductions as they have changed a lot over time and may bias distributional trends in various ways (for a very detailed description of the evolution of income tax law in France since 1914, see Piketty 2001).

We correct this in two steps, applying and updating correction factors used in Piketty (2001, 2003). We first use upgrade rates due to previous-year-tax deduction. Then, we take into account others kinds of deduction (including lump-sum deductions for wage earners) updating correction factors previously used. Compared to Piketty (2001), our only additions are: (i) we take into account the repeal of 20% deduction in 2006; (ii) we now assume that fractile P0-90 (rather than P0-100) has the same (taxable income)/(fiscal income) ratio as fractile P90-95. As a consequence the

---

<sup>43</sup> Enquête Revenus Fiscaux was produced jointly by INSEE and the tax administration from 1956 to 1990. These surveys described the socio-demographic structure of approximately 40,000 tax units along with all the information reported in their income tax returns.

approximate aggregate (taxable income)/(fiscal income) ratio is slightly smaller than what was assumed in Piketty (2001).<sup>44</sup>

Our correction factors to go from taxable income to fiscal income as well as comparisons with initial corrections used in Piketty (2001, 2003) are reported in Table TD4.<sup>45</sup> For comparison purpose, we show in Figure FD18 the difference between taxable income series and fiscal income ones. As it turns out, top income shares from fiscal income series tends to be slightly higher than taxable income ones before 1950 and lower after.

We also report upgrade rates to take into account capital gains that are estimated from total capital gains and distributional upgrade factors.<sup>46</sup>

Summary statistics and detailed series for fiscal income distribution (with and without capital gains) from 1900 to 2014 at tax unit level are provided in Tables TD10 to TD13. Figures FD10b, FD11, FD12 and FD15 show how the series change for the top income shares (from bottom 50% to top 0.001%) when the capital gains are introduced. As expected, it makes no significant difference for the bottom of the income distribution while for the top series with capital gains present slightly higher and more volatile income shares.

### Tax unit vs individual level

Next, to go from tax unit level to individual level, we use the fact that income tax tabulations also include detailed information on the numbers of married couples and of singles in each income bracket (see Table TD1). They also include the numbers of dependent children. We use the computer codes developed by Blanchet, Fournier and Piketty (2017) in order to estimate separately the distribution of fiscal income among tax units and among equal-split individuals (the initial estimates by Piketty

---

<sup>44</sup> See Table TD2 in Excel File GGP2016DINAAppendixD and do-file for average-fiscal-income adjustment. We also provide a full comparison between our new series and the old version from Piketty (2001) in Figures FD16 and FD17.

<sup>45</sup> See Piketty (2001) for a complete detail of all changes in tax law in France.

<sup>46</sup> See Table A12, Piketty (2001).

(2001, 2003) focused on tax units and did not attempt to correct for different tax unit sizes).

Summary statistics and detailed series for fiscal income distribution from 1900 to 2014 at individual level, with and without capital gains, are provided in Tables TD14 to TD17. Figures FD13a to FD14 show the direct comparison between series at the tax unit level and those at the individual level. As it turns out, series at the tax unit level tend to deliver top income shares slightly above series at the individual level. It is only for the top 10% that is it really noticeable from the 1960s and particularly after 1985. This is expected since at the tax unit level, singles and couples are ranked together although they are not directly comparable in terms of earnings. The resulting comparison between their levels of income mechanically induces a higher level of inequality. This gap is increasing from the 1960s what reflects the raising number of singles. Let us also notice that our equal-split series may under-estimate inequality because some important expenditure (such as housing expenditure) represents a heavier burden for singles than for couples.<sup>47</sup>

Finally, in table TD5, we present estimates of top taxable capital income shares in France on the long run based on income tax tabulations (1915-1998). These results are used in our companion paper (Garbinti, Goupille-Lebret, Piketty, 2016, Appendix C) to provide wealth estimates from and 1915 to 1924 and from 1965 to 1969.

---

<sup>47</sup> We do not chose a specific equivalent scale to take economies of scale into account. Had we chosen one, the resulting series would be between the estimated ones presented in Figures FD13a to FD14.