Appendix to ”Accounting for Wealth Inequality Dynamics: Methods, Estimates and Simulations for France (1800-2014)”

Bertrand Garbinti, Jonathan Goupille-Lebret, Thomas Piketty

World Inequality Lab Working papers n°2016/6

Appendix to "Accounting for Wealth Inequality Dynamics: Methods, Estimates and Simulations for France (1800-2014)"

Bertrand Garbinti, Jonathan Goupille-Lebret, Thomas Piketty

Keywords : Wealth Inequality; methods; France; tax data; national accounts; wealth surveys; wealth concentration; inequality; DINA; Distributional National Accounts;

World Inequality Lab

Accounting for Wealth Inequality Dynamics: Methods, Estimates and Simulations for France (1800-2014) Data Appendix

Bertrand Garbinti, Jonathan Goupille-Lebret, Thomas Piketty *

First Version: July 12, 2016

Last Revised: January 27, 2017

This data appendix provides methodological details and complete data series for our paper "Accounting for Wealth Inequality Dynamics: Methods, Estimates and Simulations for France (1800-2014)". It is supplemented by a set of data files and computer codes (GGP2016Wealth.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; Goupille-Lebret (Paris School of Economics, GATE-LSE): jonathan.goupille@ens.fr; Piketty (Paris School of Economics): piketty@psemail.eu. This paper presents the authors' views and should not be interpreted as reflecting those of their institutions.

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 wealth inequality dynamics.

The Appendix is organized as follows. In Appendix A, we present complete tables and additional figures of our benchmark unified wealth distributions series for the period 1800-2014. These series are obtained by combining different data sources and methods over time.

All data files, computer codes, additional series and robustness checks regarding the different methods used in our benchmark series are presented in Appendices B to D. Appendix B presents the mixed income capitalization-survey method used to estimate the distribution of wealth for the 1970-2014 period. Appendix C relates to the estate multiplier approach over the 1800-1970 period. Appendix D offers a reconciliation between estate-multiplier and income-capitalization estimates for the 1984-2010 period. Appendix E relates to the formula and simulations of long-run, steady-state level of wealth concentration. It includes the complete developments of the steady-state formula as well as additional simulations.

Finally, Appendices F to H offers a reconciliation of our benchmark series with other sources of wealth data such as household wealth surveys (Appendix F), wealth tax data (Appendix G) and rich lists (Appendix H).

This Appendix is supported by several series of Excel and PDF files as well as computer codes that contain and present our complete wealth inequality series. The directory GGP2016Wealth.zip is organized as follows. For each section of the Appendix, there is a folder called GGP2016WealthAppendixX (with X=A,…,H). 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 GGP2016WealthAppendixX.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 wealth inequality series.

Appendix A. Benchmark unified wealth distribution series (1800-2014)

Appendix A includes additional tables and figures of our benchmark unified wealth distribution series. The corresponding Excel file (GGP2016WealtAppendixA) reports wealth shares, wealth thresholds and Pareto coefficients for synthetic wealth groups (TA1) as well as for the complete 127 generalized percentiles (TA2). It includes also supplemental figures on wealth concentration in France (FA1 to FA3), on the comparison of France with the US (FA4 and FA5) and on the average wealth/national income ratio by wealth fractile (FA6 to FA8).

Appendix B. Detailed series using income capitalization method

Appendix B relates to the estimation of wealth distribution series broken down by percentile, age, gender and asset categories over the 1970-2014 period. The series are derived from a mixed method based on income capitalization and imputations from wealth surveys and housing surveys. First, we present in this section the different files included into the directory GGP2016WealthAppendixB. Then, we present the raw data sources used (income tax micro-files). Finally, we detail the different steps of the mixed income capitalization method and the robustness checks.

Section B.1. Files description

The Excel File GGP2016WealthAppendixB includes the main appendix tables and figures for detailed wealth series over the period 1970-2014. It presents tables on wealth shares and wealth composition (TB1 to TB4b) ; on the determinants of wealth inequalities by wealth group such as rates of return, real rates of capital gains, saving rates and labor income shares (TB5a to TB13) ; on other series relative to inter and intra generational wealth inequality and robustness checks (TB20 to TB22). The file reports also supplemental figures on wealth concentration (FB1 to FB2), flow returns and rates of capital gains by wealth group (FB3 to FB6), robustness checks (FB7 to FB12), wealth composition by wealth group (FB20 o FB28) and Pareto coefficients (FB30 to FB35). An index is included in the file for a complete list of tables and figures.

All the tables and figures are derived from synthetic files extracted from micro-files. These synthetic files fall into three categories. The first category of files is called "gperc" and includes for each gender and each of the 127 g-percentiles of net wealth among adults: lower wealth threshold, bottom average wealth, top average wealth and g-percentile average wealth. The second category is called "dperc". It reports wealth shares by decile[1] and gender along with a decomposition of wealth by asset categories. Finally, the last category of files reports average wealth broken down by age, gender and wealth groups. These synthetic files are included into the folder

[1] The last decile is split into 5 categories P90-95, P95-99, P99-99.5, P99.5-99.9 and P99.9-100.

DataFiles along with all computer codes and files that we use to produce homogenous wealth and income series out of income tax returns from 1970 to 2014 using capitalization method. (see ReadMe file in directory).

Section B.2. Description of data sources (income tax micro files)

As described in the core of the paper, the estimation of the wealth distribution 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[2] and the tax administration every 5 years from 1970 to 1990.[3] 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).

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.

[2] 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.

[3] 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 are now called The Tax and Social Incomes Survey (ERFS). They 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.

Finally, we also have access to income tax tabulations, which have been produced by the French Finance Ministry since the creation of income tax in France in 1914. They report the number of taxpayers and total income for a large number of income brackets. In principle, income capitalization method could also be used with tabulated income tax data broken down by income sources prior to 1970 (see Piketty 2001, Appendix B, Table B16). However, these tabulations by income categories suffer from a number of limitations[4], so that we prefer to use the income capitalization method as our benchmark method for the 1970-2014 period (when we have access to micro-files of income tax returns), and to adopt the estate multiplier method (based upon inheritance tax returns) as our benchmark method prior to 1970.[5]

Section B.2.1. Harmonization of micro files

From these different databases, we unify the concepts and names of our variables of interest which include income and demographic variables for each member of the tax unit.

- Income variables: labor and replacement incomes (wages, unemployment benefits and pensions), financial incomes (dividends, interests, life insurance income), self-employment income, real (non-imputed) rents
- Demographic variables: age of all members of the fiscal household and sex of the head of the fiscal household.[6]

As stated before, the samples of income tax returns are large (recent ones contain 500,000 fiscal units) and exhaustive at the very top. In contrast, ERF surveys are of smaller size with approximately 45,000 fiscal units. The ERF surveys are

[4] Generally speaking, the main limitation of income tax tabulations is that prior to 1985 they only cover tax units that are subject to positive income tax. Another specific limitation of 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). In contrast, inheritance tax tabulations cover the entire distribution of wealth (whether the estate is taxed), and they cover many more years prior to 1945. Full details on income tax tabulations and the way we exploit them are given in our companion paper Garbinti, Goupille-Lebret and Piketty (2017, Appendix D).

[5] As there are no inheritance tax tabulations during the period 1914-1924 and 1965-1969, we complete the missing years by using data on top capital incomes from income tax tabulations.

[6] We assume that the second member of the couple has the opposite sex of the head of the household. This assumption is relatively reasonable since couples of the same sex represents 0.6% of all couples (Buisson and Lapinte, 2013). We impute randomly the sex of the dependents. Again, this imputation has no consequences on our results as adult dependents represent less than 4% of the population, are very young (24 years old in average) and have very low income.

representative of the French population as a whole but not necessarily of the highest income earners. To check this, we compare the upper part of the income distribution from the ERF surveys with the income tax tabulations. In particular, we check the consistency of the income thresholds, the average incomes and their decomposition by income categories (rents, financial income, self-employment income, and labor income) for different income groups at the top of the distribution (P90-95, P95-99, P99-99,5, P99,5-99,9 and P99,9-100) .

The comparison between the ERF surveys and the tabulations reported in Piketty (2001) shows that average incomes by income group are almost identical but the compositions of income differ strongly for the first three surveys. Indeed, for years 1970, 1975 and 1984, labor incomes and rents are overrepresented in the top decile income of ERF surveys, while financial and self-employment incomes are underrepresented. In contrast, we find that average incomes and income compositions are identical for years 1988 and 1990 between the tabulations and the samples of income tax returns (Echantillon Léger). We then conclude that the differences we observe for the oldest ERF surveys may be due to a lack of statistical precision at the top of the income distribution. To tackle this issue, we correct the composition of the top 10% income group from the ERF surveys using the income tax tabulations.[7]

Section B.2.2. Individualization of micro files: from tax units to individual units

While the micro-files are at the tax unit level, our unit of interest is the individual level. We exploit the fact that age and gender are reported for each member of the fiscal household to fully individualize our data. More precisely, we create one observation for each member of the tax units older than 20 years old. Labor, replacement and self-employment incomes are already reported at the individual level. In contrast, financial income and rents are reported jointly for the entire fiscal household. In this case, we assume an equal split among spouses. By convention, dependents earn no

[7] More precisely, we split individuals into 6 income groups in the ERF surveys (P0-P90, P90-95, P95-99, P99-99.5, P99.5-99.9 and P99.9-100). Then we adjust each individual income and their components such as the adjusted average income for the different income groups matches those from income tax tabulations. We then follow an iterative process to adjust the income composition of the different income groups in the ERFS survey (rents, labor, financial and mixed incomes) to the appropriate one from the income tax tabulations.

financial income and no rent. Note that in France, adult children are supposed to report personally their incomes. But, when they depend on their parents' financial support, the parents can report them as dependents until they are 21 years old (at January the 1st) or 25 years old if they are pursuing an education. It is generally interesting for parents to report them because it leads to a tax reduction (thanks to the "quotient familial").

After individualizing the tax unit (and particularly the dependents), the number of young adults aged less than 25 included in our data appears to be slightly less important than in the census data. It may be due to the fact that the date of birth is not always reported for each dependent children. We then add the missing individuals aged 20-25 years with zero income, using the demographic margins by age and gender from Insee. At this stage, our micro-files are fully individualized and consistent with the French demographic structure for each year.

Section B.3. From taxable income to DINA capital income and wealth

The general idea behind the income capitalization method is to recover the distribution of wealth from the distribution of capital income flows. In its simplest form, the method relies on the assumption of fixed rates of return by asset class (see e.g. Atkinson and Harrison, 1978, and Saez and Zucman, 2016). Unfortunately, all assets do not yield taxable income and we need to adapt the capitalization method in order to impute these assets.

Here we split wealth into seven different categories of assets and corresponding capital income flows: housing assets – split into owner-occupied housing and tenant-occupied housing assets – as well as their corresponding debts, business assets, financial assets – split into equities, bonds and loans, deposits and savings accounts, and life insurance and pension funds (see Appendix A from our companion paper Table A20 for wealth and A13 for income). Four different categories of capital income are reported into the income tax returns: interests[8], dividends, self-employment

[8] Some kinds of interests are automatically taxed at "Prélèvement libératoire forfaitaire" (PLF) that is a withholding tax. Changes in the way these interests are reported into the income tax returns create artificial jumps in series for years 1995 and 1999. In the present version we do not use reported

income and tenant-occupied rental income. For these capital incomes, we estimate the corresponding stock of assets by using the income capitalization method. More specifically, we first compute the aggregate rate of return for each asset class r^i by dividing the total reported returns in the income tax returns R^i by the reported stock A^i in the national accounts. We can then obtain the stock of the different assets by dividing each capital income components reported at the individual level by the corresponding aggregate rate of return r^i.

Section B.3.1. Imputation based on household surveys

We complement the capitalization method with imputations based on household surveys (wealth and housing surveys) that present the huge advantage to report both income and wealth.[9] The imputations allow us to estimate the stock of assets that do not generate taxable income flows, namely owner-occupied housing, life insurance, and deposits (including currency and saving accounts). The imputation procedure is the following.[10]

First, in the surveys, we define groups according to three dimensions: age, financial income, and labor and replacement income. For example, we define approximately 200 groups for the imputation of owner-occupied housing asset. We first split the sample into 10 age groups (< 25 ; 25-30 ; 31-39 ; 40-49 ; 50-54 ; 55-60 ; 61-65 ; 66-70 ; 71-80 ; > 80). We then divide each age group into 4 percentile groups of financial income (P0-50 ; P50-90 ; P90-99 ; P99-100). Finally, we split again each of these 40 groups (10 age groups *4 groups of financial income) into 5 percentile groups of labor and replacement income (P0-25, P25-50, P50-75, P75-90, P90-100).

Second, for each group and each kind of asset to be imputed (owner-occupied housing, deposits, and life insurance), we both compute an extensive (the proportion of individuals holding the asset considered) and an intensive (share of the total asset owned by the group) margins. For the intensive margin of the owner-occupied

---

information about these specific interests. When computing bonds and loans, we then assume that they are proportional to the interests taxed according to the regular income tax schedule.

[9] Since 2010, the incomes reported in the wealth surveys directly come from the income tax returns.

[10] See the Stata codes in DataFiles/Capitalization method/Imputation in Appendix B for a complete and exhaustive description of the imputation method.

housing, we compute the share of total gross housing owned by the group along with the debt ratio (Debt/ Gross value of the owner-occupied housing). For instance, if 80% of the individuals in a group owns a primary residence, the total gross value of the housing asset this group owns represents 0.5% of the total value reported in the survey and their mortgage represent 50% of the gross value of their housing asset, then the extensive margin is 80%, the intensive one is 0.5% and the debt ratio is 50%.

Third, in our income tax micro files, we define groups according to the same dimensions (age, financial and labor incomes). Then, within each group, we randomly draw tax units who own the asset accordingly to the extensive margin computed for the asset and the group considered. To go back to our former example, it means that 80% of the group will be considered as owning the asset. The intensive margin is then used to impute the asset amount within the asset holders of this group. In our former example, it means that the asset-holders (who represent 80% of the considered group) will be supposed to hold globally 0.5% of the 4,484 billion euros that the gross owner-occupied housing asset represents in 2010. If the group represents 100,000 tax units, it means that each of the 80,000 tax units who own this asset will hold 0.5%*4,484 billions/80,000 that is 280,000 euros of gross owner-occupied housing. The remaining 20,000 tax units of this group won't hold any housing asset. Finally, as the debt ratio is equal to 50% in our example, the mortgage associated to the housing asset will be equal to 140,000 euros.

We should highlight that the imputations take place at the tax unit level. For life insurance and owner-occupied housing, income and stocks are then equally split among couples.[11] In contrast, deposits can be owned by all members of the tax unit (dependents and couples). For this asset, we have also computed an additional statistic for each group in the surveys corresponding to the fraction of deposits owned by the dependents. We therefore allocate the fraction of deposits owned by each member of the tax unit.

[11] The imputation at the tax unit level allows for consistency of wealth among the spouses or partners that should necessarily have the same wealth.

Finally, we estimate the different components of capital income by simply multiplying each asset by the corresponding economic rate of return (see Table A24 from Appendix A of our companion paper for the rates of return by asset computed from national accounts).

Section B.4. Robustness checks

To assess the sensitivity of our mixed capitalization method, we conducted several robustness checks (see corresponding figures and tables in the Excel file of Appendix B).

Section B.4.1. Imputation of owner-occupied housing assets

We first investigate the sensitivity of our results to the imputation of owner-occupied housing assets. In our benchmark scenario, the imputation method is based on housing surveys (1970, 1973, 1978, 1984, 1988, 1992) and wealth surveys (1992, 1998, 2004 and 2010).

Although Housing surveys are also available after 1992 (1992, 1996, 2002, 2006 and 2010), we tend to favor wealth surveys over housing surveys for the recent period. Indeed, wealth surveys include a more complete description of the different sources of income (capital, labor...), the present value of owner-occupied housing assets and the associated debts at the time of the survey.[12] This change of surveys used before or after 1992 is likely to introduce two kind of comparability issues. First, the imputation method implemented is more sophisticated when using wealth surveys instead of housing surveys.[13] Second, the distribution of owner-occupied housing and

[12] In the housing surveys, only total income is reported. Although housing occupation status is reported for all households, the value of the housing assets is asked only to home owners having bought their housing assets less than 4 years before the time of the survey. In this case, the value reported corresponds to the value of the housing assets at the time of the purchase rather than at the time of the survey. In addition, individuals are asked about their total debt rather than the mortgage associated to their housing assets.

[13] The imputations groups are defined according to three dimensions (age, financial income, and labor and replacement income) when using wealth surveys against two dimensions (age and total income) when using housing surveys. Consequently, there are 200 imputations groups when using wealth surveys against 63 groups when using housing surveys.

the associated debts is better described and more accurate in the wealth surveys than in the housing surveys.[122]

We investigate the potential comparability issues implied by the use of wealth surveys instead of housing surveys after 1992 in Figures B7 to B9 and Table B22. The first robustness check assesses the sensitivity of our results to the imputation method. In this scenario, we apply the simplified imputation procedure based on two dimensions (age and total income) to the complete period rather than only to the pre-1992 period in the baseline scenario. In the second robustness check, we use the housing surveys for the complete period 1970-2012. Figures B7 to B9 show that these changes have no significant effect on our results.

Section B.4.2. Alternative imputations of financial assets

We then investigate the sensitivity of our results to different imputation methods for financial assets. First, we impute life insurance proportionally to taxable interests and dividends rather than relying on imputation methods based on wealth surveys.[14] Second, we capitalize all financial incomes (interests from debt assets or savings accounts, life insurance income, dividends) using a unique rate of return.[15]

Note that both sensitivity checks are upper bound scenarios in terms of wealth concentration. Indeed, bonds and equities are strongly concentrated at the top of the distribution and benefit from higher returns than deposits and saving accounts that are concentrated at the bottom. Using the same capitalization factor to all financial incomes will therefore decrease the estimated value of the low-return asset and increase the estimated value of the high-return asset and therefore lead to a more important level of wealth concentration. For similar reasons, imputing life insurance proportionally to taxable interest and dividends will overestimate the level of wealth concentration.

[14] See section B.3 above for the description of the imputation method used in the baseline scenario.
[15] Interests from saving accounts and life insurance income are still imputed based on household surveys but the corresponding asset is recovered using an aggregate rate of return common to all financial income and defined as the ratio of capital income to financial assets reported in the national accounts.

Table B22 and Figures B10 to B12 report the level of wealth concentration depending on the imputations of financial assets used. The different sensitivity checks imply a slightly more important level of wealth concentration but the different trends as well as our different results and interpretations remain the same.

Section B.4.3. Capitalization of taxable interests and dividends

In our benchmark series, we jointly capitalize taxable interests and dividends to recover equities and bonds. We then reclassify them into equities or bonds proportionally to the respective importance of interests and dividends in the individual income. As it turns out, the frontier between interests and dividends reported in the income tax returns can be fuzzy and has changed over time. For instance, some capital incomes from mutual funds (FCP and OPCVM) were classified as “interests” (case TS from income tax returns) and switched to the “dividends” category (case DC from income tax returns) since 2005. This artificial change leads to a break in our fiscal series of dividends and interests that is our main motivation to capitalize them together. However, as a robustness check, we decide to test a variant where interests and dividends are capitalized separately after correcting for the artificial change occurring in 2005.[16] This allows us to test whether our conclusions are impacted by this change. Results are presented in Table B22 and Figures B10 to B12. Again, the overall impact on wealth distribution series is very limited.

Section B.5. Miscellaneous remarks

Section B.5.1. Interpolation for missing years

As described in Section B.2, micro samples of income tax returns are not available on an annual basis from 1970 to 1990. We interpolate the missing years 1971-1974, 1976-1978, 1980-1983, 1985-1987 and year 1989 by using annual aggregate series

[16] Before 2005, the case TS in the income tax returns includes interest along with capital income from mutual funds that should be reported in the case DC corresponding to the dividends. For each year before 2005, we estimate the share of capital income from mutual funds reported in case DC by using the repartition key defined as the change observed from 2004 to 2005 between the fiscal case TS and the case DC. We assume then that, between these two consecutive years and between these two fiscal cases, the essential difference comes from the change in the way households report these amounts and not from a major portfolio change. We think this is the most plausible assumption.

by asset categories from national accounts and by assuming linear trends in within-asset-class distribution. The main advantage of this interpolation method is that it allows to take into account differential shocks between assets, while remaining 100% consistent with the evolution of the aggregate stock of assets (as defined in the national accounts) over time.

We illustrate our interpolation method with the following example. Let's note Sh_{ij}^{t} the share of the asset *i* hold by the wealth group *j* at time *t* in proportion of the aggregate private wealth and $Sh\_tot_{i}^{t}$ the proportion of the asset *i* at time t into the aggregate private wealth as reported in the national accounts. For a given missing year *t* from 1970 to 1975, we have:

$$Sh_{ij}^{t} = \left(\frac{Sh_{ij}^{1970}}{Sh\_tot_{i}^{1970}} + \left[\frac{Sh_{ij}^{1975}}{Sh\_tot_{i}^{1975}} - \frac{Sh_{ij}^{1970}}{Sh\_tot_{i}^{1970}}\right] \cdot \frac{t-1970}{1975-1970}\right) \cdot Sh\_tot_{i}^{t}$$

As an alternative strategy, we also used annual income tax tabulations (broken down by income categories) and found that this makes very little difference.

Section B.5.2. Computation of saving rates, rates of return and capital gains by wealth group

Tables B5a to B8 report saving rates, rates of return and rates of capital gains by wealth group.

For each wealth group, the rates of returns are computed by weighting each asset-specific rate of return – such as reported in the national accounts[17] – by the proportion of each asset in the wealth of the group. We follow the same methodology to compute the rates of capital gains by wealth group.

[17] See Tables A23-A25 of the Appendix A of our companion paper for the rates of return and capital gain by asset computed from national accounts.

The synthetic saving rate is defined in the same way as Saez and Zucman (2016) (See discussion in the core of the paper). We first define the following transition equation:

$$W_{t+1}^{p} = (1 + a_t)(1 + q_t^p)\left[W_t^p + s_t^p \cdot Y_t^p\right]$$

With: W_t^pt, W_{t+1}^p = average wealth of group p at time t and t+1 (for instance, group p could be the top 10% wealth group)

Y_t^p = average income of group p at time t

q_t^p = average real rate of real capital gains of group p from t to t+1 (real capital gains are defined as the excess of average asset price inflation, given average portfolio composition of group p, over consumer price inflation)

a_t = average asset price inflation

s_t^p = synthetic saving rate of group p at time t

We can then compute the synthetic saving rate:

$$s_t^p = \frac{\frac{W_{t+1}^p}{(1 + a_t)(1 + q_t^p)} - W_t^p}{Y_t^p}$$

<u>*Section B.5.3. Age-Wealth profiles*</u>

We present complete series of age-wealth profiles from 1970 to 2012 in Table B21. The series are obtained by a local mean-smoothing also known as the Nadaraya-Watson estimator (1964)[18].

[18] That is in x: $\sum_{i=1}^{N} \frac{K_{h_x}(x - x_i)}{\sum_{j=1}^{N} K_{h_x}(x - x_j)}$ where $K_{h_x}(x - x_i)$ stands for the kernel function (here epanechnikov) and h_x the bandwidth around x (chosen here by the standard rule-of-thumb as implemented in the Stata procedure lpoly).

Appendix C. Detailed series using estate multiplier method

In appendix C, we present our detailed wealth distribution series obtained by applying the estate multiplier method to historical inheritance data available for France over the 1800-1970 period.

The Excel File GGPWealthAppendixC includes the main appendix tables and figures on long-term series of wealth concentration at death or among the living individuals. The file reports also supplemental tables and figures documenting step by step how wealth series are constructed starting from raw inheritance tabulations. An index is included in the file for a complete list of tables and figures. The folder DataFiles including all Stata-format and MatLab-format codes and files that we use to produce homogenous historical g-percentile wealth. The folder OtherData includes supplemental materials from other works used in this Appendix (including raw inheritance tax tabulations from Piketty 2001).

We present in this section the different inheritance data sources used and the different steps of the estate multiplier approach implemented in order to obtain the wealth distribution series over 1800-1970.

Section C.1. Inheritance data sources in France (1800-2015)

The modern inheritance tax ("droits d'enregistrement", i.e. "registration duties") was created as early as 1791 in France, as an important part of the new tax system instituted by the French Revolution. The basic features of inheritance tax law were unchanged since 1791, together with general rules to split inheritance between siblings (Code Civil, 1804).

1802-1902 Period

Before 1902, the tax administration only published aggregate statistics on the value of the estates broken down by very broad categories, e.g real (structures and buildings) and personal (furniture, businesses, stocks, bonds, etc.) assets. Hopefully, individual-level inheritance registers have been well preserved and are accessible to researchers since 1800. These registers include detailed information about assets, age, and gender, in principle for all decedents (irrespective of the level of their wealth). One needs then to return to tax registers and collect its own sample of estate tax returns to study wealth concentration.

For this period, we use inheritance tabulations based upon data collection of large individual-level micro-samples of estates collected in Paris inheritance registers (Piketty, Postel-Vinay and Rosenthal 2006) and of the provincial samples collected by Bourdieu et al (2003, 2013) in the context of the TRA survey. These raw inheritance tabulations are presented in Piketty, Postel-Vinay and Rosenthal (2006, Table A3) and are available for selected years (1807, 1817, 1827, 1837, 1847, 1857, 1867, 1877, 1887 and 1902).[19]

1902-1964 Period

In 1902, the French inheritance tax became progressive and the tax administration started compiling detailed tabulations reporting the number of decedents and amount of their wealth for a large number of inheritance brackets. These tabulations are consistent with the data collected in inheritance registers, and they are available on a quasi-annual, exhaustive national basis between 1902 and 1964 (except for the 1914-1924 sub-period). They occasionally include supplementary breakdowns by age brackets and asset categories. These raw inheritance tabulations are reported in the Excel File GGP2016WealthAppendixC (Table C3).[20]

Section C.2. Long-run series using historical inheritance data (1800-1970)

We use the inheritance tabulations in order to compute our wealth distribution series for the 1800-1970 period.[21] This work is conducted in three steps. First, the inheritance tabulations are pooled together for the complete period 1800-1964 and homogenized. Second, we apply the Pareto interpolation techniques developed by Blanchet, Fournier and Piketty (2017) to generate g-percentiles of inheritance from

[19] See Piketty, Postel-Vinay and Rosenthal (2006) and the working paper version of 2004 for more details on the data and the methodology relative to the inheritance tax tabulations.
[20] See Piketty (2001, Table J1) for a complete list of the references to the official publications from which these tabulations were copied.

inheritance tabulations. Finally, we estimate the distribution of adult wealth from the distribution of inheritance using a refined estate multiplier method.

Section C.2.1. Correction for non-filers

The inheritance tax tabulation covered only the decedents for whom an inheritance declaration was filled. Figures C1 to C3 of the Excel file report the number of decedents and inheritance declarations in France over the period 1800 to 2010. The annual number of adult decedents has generally been about 500-600 thousand throughout the 1800-2010 period in France. The annual number of inheritance declarations has generally been around 300-400 thousand (reflecting the fact that decedents with very low net wealth do not get registered via an inheritance declaration), except after the introduction of a large tax exemption in 1956, when it briefly fell to less than 100 thousand. From 1800 to 1956, the fraction of adult decedents covered by inheritance registers was therefore stable around 60%-70%.[22]

Although all inheritances were in principle subject to declaration and taxation, there has always been some tolerance for very small net wealth holders (particularly within the bottom 50% of the population, which typically owns less than 10% of aggregate wealth).[23] The tax exemption threshold introduced in 1956 led to a sharp reduction in the number of declarations (although in principle declaration was still compulsory). The threshold was under-indexed in the following decades, and the fraction of tax filers gradually returned to earlier levels.

[22] Note that TRA samples clearly show a slow decline of the proportion of adult decedents with inheritance declarations during the 19th century (from about 70% in the early 19th century to about 60% by the end of the century), and that this seems strongly related to urbanization (property is less widespread in cities than in the country side). See Bourdieu et al (2003, figure 3; 2013, figure 11, p.147, tables 12-13, p.183). See also the very low proportion of inheritance declarations in Paris (Figure FC3 of the Excel file and Piketty, Postel-Vinay and Rosenthal (2006, 2014)). There is however some uncertainty about the exact proportion of inheritance declarations at the national level during the 19th century, due in particular to the limited size of the TRA samples. Also note that TRA samples and national tabulations are not entirely consistent on this issue. E.g. in 1902-1910 TRA samples find inheritance declarations for 56% of adult decedents, against 66% according to national tabulations. One possible interpretation would be that national tabulations include multiple individual declarations (when new information is added to the main declaration). See Bourdieu et al (2003, table 2 and ensuing discussion). If this was the case, then this would imply that national tabulations tend to under-estimate wealth concentration (as multiple declarations are more widespread for large estates). Another interpretation is that the TRA sample is not entirely representative.

[23] See the Appendix of Piketty (2010) for a complete discussion of the estate tax data in France and tax filling requirement.

The only correction we made to the inheritance tax tabulations was to add non-filers and their corresponding wealth. The corrected tabulations are presented in the Excel File GGP2016WealthAppendixC (Tables C4 for 1902-1964 and C5 for 1807-1887). These tables cover all adult decedents (filers and non-filers) and assume that the ratio between non-filers and filers average wealth (znf) is equal to 2%.[24]

Table C6 reports the homogenized corrected inheritance tabulations 1807-1964 used for generalized Pareto simulations. It depicts wealth thresholds and Pareto coefficients by percentiles (P10, P50, P90, P95, P99, P99.5, P99.9 and P99.99). In order to model explicitly for the functional form of the wealth distribution at the bottom (see below for the description of the generalized Pareto simulations), we assume that the wealth threshold of the first decile (P10) is set to fit post-1970 P10/P50 ratio pattern.[25] This assumption has no impact on wealth inequality as the wealth share of the first decile has always been very low (less than 0.2% of total wealth).

Section C.2.2. Generating g-percentiles of inheritance from inheritance tabulations

We apply the generalized, non-parametric Pareto interpolation techniques developed by Blanchet, Fournier and Piketty (2017) to the corrected inheritance tabulations (Table C6) in order to estimate the complete distribution of wealth at death among decedents over 1807-1964 period. Table C10 reports summary statistics for the distribution of wealth at death, while Table C11 presents the detailed series by g-percentiles (from P1 to P99.999). All the computer codes and files in MatLab and Stata formats are gathered in the folder DataFiles of the Appendix C along with ReadMe files presenting the different programs and output tables.

Note that we made two additional adjustments to obtain the complete series of wealth at death during the 1807-1964 period. First, there is a discrepancy between fiscal and economic flows of wealth at death due to legally tax-exempt assets and tax evasion (see Table C2). The series were anchored to economic wealth at death

[24] We also perform sensitivity checks with different values for znf and found that the impact on wealth inequality was relatively small.
[25] P10 represents 7.7% of P50 in 1970 and ranges from 0.6% to 1.3% of average wealth over 1807-1964 period. See formulas in Tables C6 and C7 of the Excel file for more details.

instead of fiscal wealth at death by simply multiplying each amount by the ratio between economic and fiscal wealth at death.[26] Second, we estimate the distribution of wealth for the years not covered by inheritance tax tabulation (1914-1924 and 1965-1969) using data on top capital incomes from income tax tabulations. Table C7bis reports the index factors used for these periods.

Section C.2.3. Estimating the distribution of adult wealth from the distribution of inheritance

Table C12 reports summary statistics for the distribution of wealth among adults, while Table C13 presents the detailed series by g-percentiles. All details of the methodology used in order to estimate the distribution of adult wealth from the distribution of inheritance over the 1807-1964 period are given in the computer codes and do-files. We use the same ratios as those used in Piketty 2014 (see xls files Piketty2014Chapter10TablesFigures and Piketty2001TaxAnnexeJNov2015). These ratios are taken from Piketty, Postel-Vinay and Rosenthal (2006) and are based upon the application of the estate multiplier method to the micro-level files collected in Parisian and national (TRA sample) inheritance archives between 1807 and 1964, as well as the national tabulations by inheritance and age brackets published by the Finance Ministry between 1902 and 1964.

It should also be noted that the ratios living/decedents obtained for top wealth shares over the 1807-1964 period are typically larger than one (i.e. there is somewhat more inequality when we include all living age groups together than when we only look at decedents), but not that much larger than one (with ratios typically around 1.01 for top 10% shares and 1.05 for top 1% wealth shares). Note also that Piketty (2014) also uses the series from Piketty-Postel-Vinay-Rosenthal (2006) for 1984 and 1994 (based upon DMTG files and estate multiplier method) and from Landais-Piketty-Saez (2011) for year 2010 (based upon income capitalization method). The series used in the present paper for the entire 1970-2014 are based upon a mixed income capitalization-wealth survey method and should be viewed as more precise.

[26] We do not try to correct the distribution using differential factors by level of wealth. In absence of any external sources available at the micro level, it seems more reasonable to make a neutral assumption in terms of distribution.

Appendix D. Reconciliation between the two methods (income capitalization and estate multiplier)

In our benchmark series, we combine wealth series derived from estate multiplier approach using inheritance tax data from 1800 to 1970 with those obtained from a mixed income capitalization-survey method using income tax returns and wealth surveys from 1970 to 2014. The choice of the capitalization method over the estate multiplier approach for the recent period is led by the fact that the former allows to draw a more complete picture of wealth inequality. Indeed, the capitalization method allows to estimate the joint distribution of income and wealth broken down by age, gender and asset categories. However, we show in the present Appendix D that wealth series derived from the capitalization method and the estate multiplier approach depict similar results.

Section D.1. Files description

The Excel File GGPWealthAppendixD includes the main appendix tables and figures for wealth series derived from the estate multiplier approach from 1984 to 2010. Tables D1 to D5 report the demographic parameters (population, number of decedents and mortality rate) and differential mortality rates (which in general could vary by gender, age, time and wealth percentiles) used to apply the estate multiplier approach. Wealth inequalities at death and for the overall population are also reported (Tables D6 to D8) as well as sensitivity analysis (Tables D9 to D10, see below). The file reports also supplemental figures on wealth concentration at death or for the overall population along with a comparison of wealth series derived from the capitalization method or the estate multiplier approach. An index is included in the file for a complete list of tables and figures.

The folder DataFiles includes all codes and files that we use to produce homogenous wealth series out of national micro-samples of inheritance tax returns (DMTG microfiles) covering the years 1984, 1987, 1994, 2000, 2006 and 2010 (see ReadMe file in directory).

Section D.2. Estate multiplier approach: General method

Section D.2.1. Mortality rates

The estate multiplier approach allows to recover the wealth distribution for the overall population from the distribution of wealth at death. This approach stems from the simple definition of the mortality rate m=D/L where m stands for the mortality rate, D the number of decedents and L the number of living individuals. From the number of decedents, it is thus easy to compute the number of living: L = D/m. We can therefore recover the wealth distribution among the living from micro samples of inheritance tax returns by simply reweighting each decedent by the inverse of the mortality rate. However, this method relies on the assumption that death can be seen as an exogenous event.

It is thus of concern to know to what extent death is purely exogenous or if it is correlated with some characteristics. Indeed, the probability of dying turns out to depend on socio-demographic features. For instance, it is lower for women, it decreases with wealth or over time and increases with age. It is not a problem as long as those determinants can be taken into account through mortality rates differentiated by gender, age, time and wealth. The underlying assumption behind the validity of these differential mortality rates is then weaker: it is only required that death is exogenous conditionally on those variables, i.e. within each class group defined by gender, age, time and level of wealth.

We compute differential mortality rates using mortality tables provided by Blanpain (2016).[27] These tables report mortality rates for different periods (1976-1984, 1983-1991, 1991-1999, 2000-2008 and 2009-2013), by gender, age and Socio-Occupational Categories (PCS).[28] We apply the mortality rates computed for 1983-

[27] "Les inégalités sociales face à la mort – Tables de mortalité par catégorie sociale et diplôme", N. Blanpain, 2016 (Insee Résultats)

[28] There are 7 PCS : white-collar workers (employés) ; blue-collar workers (ouvriers) ; managers and intellectual professions (cadres et professions intellectuelles supérieures) ; intermediate occupations (professions intermédiaires) ; inactives (inactifs) ; craftsmen, traders and company mangers (artisans, commerçants et chefs d'entreprise).

1991 to the DMTG microfiles 1984 and 1987, for 1991-1999 to the DMTG microfile 1994, for 2000-2008 to the DMTG microfiles 2000 and 2006 and for 2009-2013 to the DMTG micro files 2010. We define 7 age groups (< 40; 40-49; 50-59; 60-69; 70-79; 80-89; > 90). We define three wealth groups (top 10%, middle 40% and bottom 50%) based on the 7 Socio-Occupational Categories. For the top 10%, we use the mortality rate corresponding to executives and intellectual professions ("professions intellectuelles supérieures"). For the middle 40 %, we compute the mortality rate corresponding to intermediate professions ("professions intermédiaires"). Finally, the mortality rate for the bottom 50% is obtained such as: [29]

$$m = 0.1 * \ m_{top} + \ 0.4 * \ m_{mid} + \ 0.5 * \ m_{bot}$$

Where m, m_{top}, m_{mid}, m_{bot} are respectively the average mortality rate, the mortality rate of the top 10%, the middle 40% and the bottom 50% for a given decennial age, time and gender group. Table D5 reports all the differential mortality rates computed and compare them with those used by Piketty (2011).[30]

A last concern with the estate multiplier approach rises from the fact that individuals may be able to anticipate the date of their death several years in advanced and accordingly change their level of wealth. For instance, with the dramatic increase in life expectancy, old individuals may have more time to plan for the disposition of their estate and may give *inter-vivos* gifts to decrease their inheritance tax liabilities. They may also have faced health problems, dependency and medical expenditures several years before dying due to terminal illness. In these cases, wealth of the decedents would be lower than the wealth of the living individuals from a similar gender and age. The estate multiplier approach would therefore under-estimate the wealth among the living. If this bias is time invariant between and within each class group, then the wealth shares will not be biased (even though the level of wealth will).

---

[29] With the same notations as above:
$m.L = D = D_{top} + D_{mid} \ + D_{bot} \ = m_{top} \cdot L_{top} \ + \ m_{mid} \cdot L_{mid} + \ m_{bot} \cdot L_{bot}$
and so $m = m_{top} \cdot \frac{L_{top}}{L} + \ m_{mid} \cdot \frac{L_{mid}}{L} + \ m_{bot} \cdot \frac{L_{bot}}{L} = \ 0.1 * \ m_{top} + \ 0.4 * \ m_{mid} + \ 0.5 * \ m_{bot}$

[30] In these tables we provide "relative mortality" that corresponds to the ratio of the mortality rate of the group divided by the mortality rate of the population. The mortality rates used by Piketty (2011) were only varying with decennial age and two wealth groups (bottom 50% and top 50%).

Some variability may eventually appear from one year to another. They are inherent to the estate multiplier approach. Indeed, for each year, there are 500,000 decedents in France. The accidental death of a young (less than 60 years old) billionaire may then lead to a strong increase in the top wealth share because the wealth of this billionaire would be multiplied by a high coefficient due to the low death probability of young individuals.[31]

Section D.2.2. Accounting for differential mortality rates by wealth group

The simple definition of the mortality rate ($L = \frac{D}{m}$) presented above can be extended to take into account differential mortality rates. For instance, when the living population is split equally into two wealth groups, the number of living individuals of each group is given by:

$$L^P = \frac{D^P}{m^P} \text{ and } L^R = \frac{D^R}{m^R},$$

where the subscript P refers to the Poor and R to the Rich.

However, if the Rich have a lower mortality rate (ie they live longer than the Poor), then the share of rich decedents (within the total number of decedents) will be less than a half. It is thus not completely obvious to deduct the share of decedents of one group directly from the share represented by this group within the living population. We detail below how to compute the different shares of decedents by group in a wider framework where the living population is split into 3 groups (the Rich, the Middle class and the Poor). We note:

- m^R, m^M and m^P are respectively the mortality rates of the Rich, the Middle class and the Poor group.
- D the total number of decedents, D^R, D^M and D^P are respectively the number of decedents from the Rich, the Middle class and the Poor group.

[31] For instance in 2011, Saez and Zucman (201) document a huge increase in the top 1% due to Steve Jobs' death. On French data, we face a similar problem for year 2010. We correct this by replacing the age of this billionaire (43 years old) by the average age at death.

- L the total number of living individuals, L^R, L^M and L^P are respectively the number of living individuals from the Rich, the Middle class and the Poor group.

Then,

$$m^R = \frac{D^R}{L^R},\ m^M = \frac{D^M}{L^M} \text{ and } m^P = \frac{D^P}{L^P}$$

If the living population is split into three parts such as the Rich represent 10% of the total population (ie $L^R = 0.1L$), the Middle class 40% ($L^M = 0.4L$) and the Poor 50% ($L^P = 0.5L$), then:

$$D = D^R + D^M + D^P$$

$$D = D^R + \frac{m^M}{L^M} + \frac{m^P}{L^P}$$

$$D = D^R + 0.4m^M L + 0.5m^P L$$

$$D = D^R + \frac{0.4}{0.1} m^M L^R + \frac{0.5}{0.1} m^P L^R$$

$$D = D^R + 4\frac{m^M}{m^R} D^R + 5\frac{m^P}{m^R} D^R$$

$$D = (1 + 4\frac{m^M}{m^R} + 5\frac{m^P}{m^R})D^R$$

From this, it is straightforward to compute the number of decedents for each group:

$$D^R = \frac{m^R}{m^R + 4m^M + 5m^P} D$$

$$D^M = m^M L^M = m^M \frac{0.4}{0.1} L^R = 4\frac{m^R}{m^R} D^R$$

$$D^P = 5\frac{m^P}{m^R} D^R$$

It is easy to generalize the simple case presented above to a population split into N groups, such as $L^j = p^j L$ (with p^j the proportion of living individuals from group j). Taking L^1 as a reference, we can write:

$$D = D^1 + \sum_{j>1}^{N} D^j$$

$$D = D^1 + \sum_{j>1}^{N} m^j L^j$$

$$D = D^1 + \sum_{j>1}^{N} m^j \frac{p^j}{p^1} L^1$$

$$D = D^1 (1 + \sum_{j>1}^{N} \frac{m^j}{m^1} \frac{p^j}{p^1})$$

From this, we compute the number of decedents of each group:

$$D^1 = D \frac{m^1}{m^1 + \sum_{j>1}^{N} \frac{p^j}{p^1} m^j}$$

$$D^j = \frac{m^j}{m^1} \frac{p^j}{p^1} D^1$$

Section D.3. Description of data sources and corrections

We use micro-samples of inheritance tax returns to apply the estate multiplier approach to the 1984-2010 period. These micro-files, called "DMTG files"[32], have been produced by the French Finance Ministry every 6-7 years since 1977. We have access to the six existing waves of the files: 1984, 1987, 1994, 2000, 2006 and 2010.[33] Each file contains between 3,000 and 5,000 individual estate tax returns (as compared to a total of about 300,000 estate tax returns filed each year). Hopefully, the wealthiest decedents are heavily oversampled, so that DMTG files can be representative of the very top of the distribution. The files include all variables reported in the estate tax returns, and in particular detailed information on the value of the estate broken down by asset categories along with socio-demographic characteristics of the decedent and his/her heirs and the share of the estate accruing to each heir.[34]

Even though these micro-files provide very rich information on intergenerational wealth transmission, they present three potential drawbacks.

[32]"DMTG" stands for "Droits de mutation à titre gratuit", which is the official name of the estate tax in France.

[33] The 1977 DMTG file has not been archived and is no longer available.

[34] See Piketty (2011), Appendix B for a complete presentation of the tax data related to estate taxation in France.

The first one is due to the fact that an estate tax return is not established after each death, mainly because of (implicit or explicit) tax filling thresholds.[35] As a direct consequence, the lowest estates are not included in the sample.[36] For instance, the weighted number of decedents in the 2000 micro-file represents 66% of the total number of decedents, and 50% in 1984 and 1987. It is then necessary to adjust the data to include all decedents and to impute them the appropriate level of wealth.

The second one has to do with tax-exempt assets, i.e. assets partially or fully exempt from tax and therefore not entirely reported in estate tax returns. For instance, most life insurance assets do not have to be reported in the estate tax return since they benefit from a specific tax treatment at death. Life insurance has then to be imputed.

Finally, since estate tax returns are established for taxation purpose, we cannot rule out some illegal under-reporting in order to decrease the inheritance tax liabilities. We then rescale the reported assets of the living obtained after implementing the estate multiplier approach such as to match the wealth composition from national accounts.

We detail these two points below.

Section D.3.1. Correction for non-filers

As mentioned above, the lowest estates are not always subject to an estate tax return and are therefore not included into the DMTG files. For instance, the number of estate tax returns represented 50% of the number of adult decedents in 1984 and 1987, 60% in 1994, 66% in 2000 and 2006 and 58% in 2010. Therefore, the proportion of missing decedents in the DMTG files varies from 34 to 50% over time. We proceed in two steps to correct for non-filers.

[35] A tax filling threshold of 10,000 euros was introduced in 2004 for spouses and children heirs that was raised to 50,000 euros in 2006 in the absence of inter-vivos gifts. Before 2004, the tax administration tolerates that small estates may not be reported if they were not taxable. See Piketty (2011), Appendix B for a complete discussion on non-filers and tax filling thresholds.

[36] There may exist some rare cases where estates below the tax filling threshold are reported to the tax administration.

First, we add missing decedents to the DMTG files to get a fully representative sample of the demographic structure of the French decedents over time. The adjustment consists in adding non-filers at death such as our new files match the demographic structure of the decedents (by age and sex) provided by Insee. This first step simply consists in adding decedents with zero wealth but with appropriate age and gender in order to replicate the French structure of decedents. We can then apply the estate multiplier method to get a representative sample of the living population (according to the estate multiplier approach described above). In this sample, the living individuals corresponding to the non-filers with zero wealth represent 35 to 45% of the population.

The wealth of the non-filers has always been very low. However, we need to impute it because the proportion of non-filers has varied over time. Had we not imputed non-filer wealth, the wealth share accruing to the bottom 50% would have changed due to the evolution of the proportion of non-filers over time.[37] In a second step, we then impute the wealth of the bottom 50% of the wealth distribution using a uniform law and assuming that the average bottom 50% wealth level is the same as estimated with the capitalization method. Ideally, it would be better to model explicitly the functional form of the wealth distribution at the bottom. As we are only interested in the overall bottom 50% wealth share, such an explicit modeling is however far beyond the scope of this work. Moreover, we tried several alternative assumptions and we found that their impact on wealth concentration was relatively small.

Section D.3.2. Correction for tax-exempt assets

The second adjustment that has to be made is relative to tax-exempt assets. In theory, all assets transmitted at death have to be reported at their market value whether they are taxable or not. Several factors may nevertheless explain an underreporting of the assets.

[37] For instance, the living individuals corresponding to the non-filers with zero wealth represent 46% of the overall population in 1987 against 34% in 2006. Without any imputation of non-filer wealth, the bottom 50% wealth share would have increased from 0.6% in 1987 to 4% in 2006, while in reality the bottom 50% wealth share was slightly decreasing from 9% in 1987 to 7% in 2006.

First, the fiscal administration allows non-taxable assets not to be reported since they are not subject to the estate tax. This is mainly the case for life insurance assets.[38] When comparing the total amount of life insurance transmitted at death (data from the French Federation of Insurance Companies) with the reported amount in the DMTG micro files, we notice that approximately 90% (and 100% before 1992) of life insurance assets are not reported into the estate tax returns. This is particularly problematic as the total amount of life insurance transmitted at death has dramatically increased over time from 11% to 31% of the fiscal bequest flow between 1984 and 2006. Similarly, other assets benefit from various exemptions and special tax rebates and only their taxable fraction is reported in the estate tax (in a smaller proportion than life insurance assets).[39]

Second, assets may be undervalued when there is a steep increase in housing or stock market prices. Indeed, individuals are likely to not fully take into account this increase in the way they value their assets. As a direct consequence, the reported value of the assets does not fully account for the recent changes in prices and may then differ from their current market value.

Eventually, we cannot rule out some systematic (and illegal) under-reporting of the estate by heirs in order to minimize tax liabilities.

Section D.3.3. Correction for life insurance assets

We impute life insurance to the top 50% of the wealth distribution using wealth surveys - as the bottom 50% wealth and its component (including life insurance assets) was already imputed during the previous step (see above). The total level of life insurance imputed is equal to the aggregate stock of life insurance assets

[38] Before 1992, life insurance assets transmitted at death were fully exempt from taxation at death. Since 1992, for life insurance policies subscribed after 11/20/1991, only contributions made after age 70 and above 30,500 euros are subject to estate tax (not the corresponding interest). Since 1998, all contributions not subject to estate tax are taxed at death at a specific tax rate of 20% after a tax exemption of 152,500 euros. However, this special tax is independent from the general estate tax and the corresponding asset values are not reported in estate tax returns.

[39] The primary residence of the decedent benefits from a 20% rebate on market value when the surviving spouse or one of the children lived in it with the decedent. Family firm, specific rural assets or the first intergenerational transmission of all real assets built between 1947 and 1973 benefits also from specific exemptions. See Piketty (2011) for a detailed description of tax-exempt assets over time.

reported in the national accounts minus the amount owned by the bottom 50% as assessed with the capitalization method.

The spirit of the imputation procedure is the same as that used in the mixed income capitalization-survey method.[40]

First, in the surveys, we define groups according to three dimensions: age, non-life insurance-financial assets (stocks, bonds and savings accounts) and net tangible assets (housing and business assets minus liabilities). For example, we define approximately 150 groups. We first split the sample into 5 age groups (< 40; 40-49; 50-59; 60-69; ≥ 70). We then divide each age group into 6 groups of non-life insurance-financial assets (P0-50, P50-75, P75-90, P90-95, P95-99, P99-100). Finally, we split again each of these 30 groups (5 age groups*6 groups of non-life insurance-financial assets) into 5 percentile groups of tangible assets (P0-25, P25-50, P50-75, P75-90, P90-100).

Second, for each of these final groups, we compute in the wealth surveys an extensive margin (the probability of owning life insurance assets within the group) and an extensive one (the share of the total amount of life assurance that is owned by the group).

Third, in our DMTG files, we define groups according to the same distinction (age, non-life insurance-financial assets and tangible assets). Then, within each group, we randomly draw individuals who own the asset accordingly to the extensive margin computed for this group. The intensive margin is then used to impute the asset amount within the asset holders of this group.

Section D.3.4. Correction for underreporting of other assets

At this stage, the micro files are updated to include the aggregate stock of life insurance assets such as reported in the national accounts and non-filer wealth (bottom 50% wealth). Other assets of the top 50% are then rescaled asset by asset to the corresponding stock reported into the national accounts in order to take into account potential under-reporting.

[40] See Section B.3 for an example of the imputation procedure.

Section D.3.5. Results

Table D7 and Figures D1 and D2 show the comparison between the series on wealth inequality derived either by the capitalization method or by the estate multiplier approach. Both series depict the same level and evolution of wealth inequality.

We now present sensitivity analysis of wealth inequality using the estate multiplier approach.

Table D9 and Figure D3 show that wealth inequality in France is not affected by the choice of differential mortality parameters. The results are quite similar whether we apply the estate multiplier approach without differential mortality parameters, with differential mortality parameters varying by age and level of wealth (top 50% vs bottom 50%) or with our benchmark differential mortality parameters varying by age, level of wealth (top 10% vs middle 40% vs bottom 50%) and time periods (1976-1983, 1983-1991, 1991-1999, 2000-2008, 2009-2013).[41]

Table D10 and Figure D4 present the evolution of wealth inequality before and after the adjustments made (correction for non-filers and correction for tax-exempt assets). Adding non-tax filer wealth decreases slightly the top 1% wealth share. On the contrary, the correction for tax-exempt assets increases moderately the top 1% wealth share as financial assets (and particularly life insurance assets), which are much more concentrated at the top, represent the major source of tax-exemption. Taken together, the adjustments have only a very moderate impact on the top 1% wealth share, denoting the robustness of our series.

[41] Table D5 presents the different differential mortality parameters used in our estate multiplier approach.

Appendix E. Simulations

In the Section 6 of the main paper, we present different simulation exercises derived from the transition equation. We present in this appendix the details of the computations made. The first section is dedicated to the simulations of top 1% wealth share over the period 1970-2014 (Figures 12 and 13 of the core paper). The second section reports the technical details relative to the steady-state formula for wealth concentration along with two simulations of steady-state trajectories (Figure 17 of the core paper). We describe also alternative simulations of steady-state trajectories reported in the Excel File GGP2016WealthAppendixE.

Section E.1. Simulating the evolution of top 1% wealth share (1970-2014)

The purpose of the simulations is to investigate the impact of some key forces at play (capital gains and saving rate) during 1970-2014 period. We start from the accumulation equation of asset A from wealth group i at time t + 1:

$$A_{t+1}^{i} = (1+p_t)\left(1+q_{t,A}\right)\left(1+\ s_{t,A}^{i}\right)A_t^{i}$$
$$\Rightarrow A_T^{i} =\ \prod_{t=t0+1}^{t=T}(1+p_t)\left(1+q_{t,A}\right)\left(1+\ s_{t,A}^{i}\right)\ A_{t0}^{i}$$

Where s_A is the saving-induced asset growth rate (in % of asset A), p is the inflation rate and q is the real rate of capital gain.

The first simple simulation exercise consists of replacing the time-varying rates of real capital gains $q_{t,A}$ by constant capital gains $\bar{q}_A$ (namely by the average structural increase and decrease of the various asset prices over the 1970-2014 period) using the following equation:

$$A_T^{i} =\ \prod_{t=t0+1}^{t=T}(1+p_t)(1+\bar{q}_A)\left(1+\ s_{t,A}^{i}\right)\ A_{t0}^{i}$$

The second simple simulation exercise consists of replacing both the time-varying rates of real capital gains and the saving-induced asset growth rate by their averages over the period 1970-2014. The idea is to investigate the structural increase of capital gain and wealth accumulation stripped of large short run fluctuations. This is done by applying the following equation:[42]

$$A_T^i = \prod_{t=t0+1}^{t=T} (1+p_t)(1+\bar{q}_A)(1+\bar{s}_A)\frac{(1+s_{t,A}^i)}{(1+s_{t,A})}\ A_{t0}^i$$

Where $\bar{s}_A$ stands for the average savings-induced asset growth rates and $s_{t,A}$ the time-varying savings-induced asset growth rates.

In this equation, we assume that each asset A grows at the same rate $(1+\bar{q}_A)(1+\bar{s}_A)$, i.e. its average growth rate over the period 1970-2014 corresponding to the product of the average savings-induced asset growth rates and the average rate of real capital gains. However, we want our simulation takes into account structural changes in wealth accumulation behavior. To do this, we allow the savings-induced asset growth rate to vary along the wealth distribution and over time by weighting the aggregate average savings-induced asset growth by the ratio $\frac{(1+s_{t,A}^i)}{(1+s_{t,A})}$, i.e. the ratio of the savings-induced asset growth of group i at time t by the aggregate savings-induced asset growth at time t.

Figure 12 (presented in the main paper) reports the results of the simulations when savings-induced asset growth rates and/or capital gains are replaced by their averages over the period 1970-2014. By construction, all simulated series end up in 2014 at the same inequality level as the observed series. Figure 13 reports the results of the simulation when the rates of capital gains and savings-induced asset growth rates are replaced by their averages over the 1970-2000 period, i.e. over the period ending before the housing boom of the 2000s.

[42] We compute for each year an asset corrected for the short run fluctuations and whose capital gain and savings-induced asset growth rate evolutions are the average ones observed over the period. We then rescale the asset evolution within each wealth group using this counterfactual evolution. It simply implies to use as capitalization factor the ratio obtained dividing the total flow of returns of each asset by its corrected amount (instead of using the amount such as observed in the national accounts).

Section E.2. Detailed equations on the steady-state formula

Section E.2.1. Steady-state formula

We start from a simple accounting equation to describe wealth accumulation between period t and t+1:

$$W_{t+1} = (1+p_t)(1+q_t)\left(W_t + s_t\left(Y_{L_t} + r_t W_t\right)\right)$$

Where p is the inflation rate, q is the real rate of capital gain, W is the aggregate wealth, s is the saving rate (in % of pretax income), Y_L is the labor income and r is the rate of return.

At t+1, the wealth share of wealth group p is:

$$sh^p_{W_{t+1}} = \frac{W^p_{t+1}}{W_{t+1}} = \frac{1+q^p_t}{1+q_t}\frac{W^p_t + s^p_t(Y^p_{L_t} + r^p_t W^p_t)}{W_t + s_t(Y_{L_t} + r_t W_t)}$$

Using the fact that $\frac{W_t}{Y_{L_t}} = \frac{W_t}{Y_t}\frac{Y_t}{Y_{L_t}} = \frac{\beta_t}{1-\alpha_t}$ and dividing by W each term of the fraction, we obtain:

$$sh^p_{W_{t+1}} = \frac{1+q^p_t}{1+q_t}\frac{sh^p_{W_t}(1+s^p_t r^p_t) + s^p_t(\frac{1-\alpha_t}{\beta_t})sh^p_{Y_{L_t}}}{1 + s_t r_t + \ s_t(\frac{1-\alpha_t}{\beta_t})}$$

Where α is the capital share and β is the wealth-income ratio.

Assume the relative capital gain channel disappears, i.e. all asset prices rise at the same rate in the long run (which must happen at some point, otherwise there will be only one asset left), and this rate is the same as consumer price inflation (otherwise wealth-income ratio would go to infinity).

At the steady state, $sh^p_{W_t} = sh^p_W$ and is equal to:

$$sh^p_W = \frac{s^p(\frac{1-\alpha}{\beta})}{s\left(r + \frac{1-\alpha}{\beta}\right) - s^p . r^p} sh^p_{Y_L} \quad (1)$$

Using $\beta = \frac{s}{g}$ and $\alpha = \frac{sr}{g}$,[43] the steady-state formula (1) can alternatively be computed as:

$$sh_W^p = (1 + \frac{s^p r^p - sr}{g - s^p r^p})\frac{s^p}{s} sh_{Y_L}^p \quad (2)$$

With

$$\begin{cases} r = r^p . sh_W^p + r^{1-p}\left(1 - sh_W^p\right) \\ s = s^p . sh_Y^p + s^{1-p}\left(1 - sh_Y^p\right) \\ sh_Y^p = \frac{s}{g}\left(r^p . sh_W^p - r . sh_{YL}^p\right) + sh_{YL}^p \ ^{44} \end{cases} \quad (3)$$

The resolution of the system *(2)* and *(3)* gives:

$$sh_W^p = \frac{s^p . sh_{Y_L}^p \ \ (g - s^{1-p} r^{1-p})}{(g - s^p r^p)\left[s^{1-p}\left(1 - sh_{Y_L}^p\right) + s^p . sh_{Y_L}^p\right] + s^p . sh_{Y_L}^p \ \ (s^p r^p - s^{1-p} r^{1-p})} \ (4)$$

The steady-state top 10% wealth shares derived from (4) are presented in Tables E1 and E2. Table E1 shows the top 10% steady-state wealth share based on the historical values of the parameters over different time periods (1970-1984 or 1984-2014). Table E2 reports six hypothetic scenarios illustrating the relative importance of the different parameters (growth rate, inequality of rates of return and saving rates) on the steady-state wealth shares.

Section E.2.2. Steady-state trajectories

Tables E3 and E4 of the Appendix E show the steady-state wealth share trajectories until 2150. These simulations are based on the historical values of the parameters g, Y_L^p, r^p, s^p, s^{1-p}, r^{1-p} computed over 1970-1984 or 1984-2014 periods and that we assume to be constant over time. Figure 17 of the main paper is derived from these tables. We now present the different equations used to estimate these trajectories.

[43] Using $\beta = \frac{W}{Y} = \frac{s}{g}$, $\alpha = \frac{rW}{Y} = r.\beta = r\frac{s}{g}$ at the steady state.

[44] sh_Y^p is obtained by replacing $\beta = \frac{s}{g}$ and $\alpha = r\frac{s}{g}$ in $sh_Y^p = r^p . sh_w^p . \beta + sh_{YL}^p(1 - \alpha)$

The wealth share of the group p (e.g top 10%) at time t+1 can be derived from the following equation:

$$sh^{p}_{wt+1} = \frac{W^{p}_{t+1}}{W_{t+1}} = \frac{W^{p}_{t} + s^{p}_{t}.Y^{p}_{t}}{W_{t} + s_{t}.Y_{t}}$$

By dividing by Y_t each member of the fraction and using the fact that $\frac{W^p_t}{Y_t} = \beta_t.sh^p_{Wt}$, we obtain:

$$sh^{p}_{wt+1} = \frac{\beta_t.sh^{p}_{Wt} + s^{p}_{t}.sh^{P}_{Yt}}{\beta_t + s_t}$$

As $W_{t+1} = W_t + S_t$, by dividing successively each member of the equation by Y_t and Y_{t+1} and using the fact that $g_t = \frac{Y_{t+1}}{Y_t}$, $\beta_t = \frac{W_t}{Y_t}$ and $s_t = \frac{S_t}{Y_t}$, the wealth-income ratio at time t+1 is equal to:

$$\beta_{t+1} = \frac{\beta_t + s_t}{1 + g_t}$$

The aggregate rate of return and the capital share at time t+1 are respectively equal to:

$$r_{t+1} = r^{p}_{t+1} \cdot sh^{p}_{Wt} + r^{1-p}_{t+1}\left(1 - sh^{p}_{Wt+1}\right),\ \alpha_{t+1} = r_{t+1}.\beta_{t+1}$$

The expression of sh^P_{Yt+1} is given by:

$$sh^{p}_{Yt+1} = \frac{r^{p}_{t+1}W^{p}_{t+1} + Y^{p}_{L_{t+1}}}{r_{t+1}W_t + Y_{L_{t+1}}}$$

Dividing each member of the equation by W^p_{t+1} and using the fact that $\frac{Y_{Lt}}{W_t} = \frac{1-\alpha_t}{\beta_t}$, $\frac{Y^p_{Lt}}{W_t} = \frac{(1-\alpha_t)sh^P_{YLt}}{\beta_t}$ and $\alpha_t = r_t.\beta_t$, we obtain finally:

$$sh^{p}_{Yt+1} = \beta_{t+1}.r^{p}_{t+1}.sh^{p}_{Wt} + (1 - \alpha_{t+1}).sh^{P}_{YLt+1}$$

Finally, we also need to compute the aggregate saving rate at time t+1 which is equal to:

$$s_{t+1} = s^{p}_{t+1} \cdot sh^{p}_{Yt+1} + s^{1-p}_{t+1}(1 - sh^{p}_{Yt+1})$$

Section E.2.3. Steady-state formula with capital depreciation or appreciation

We now extend the formula assuming some exogenous rate of capital depreciation or appreciation q. q<0 corresponds to depreciation and q>0 to appreciation. Note that the rate of capital gains could be endogenized via multi-sector growth models with differing rates of technical progress. Here, we take it as given for simplicity.

In this new framework, the steady-state wealth share formula remains equal to (1) as $q_t^p = q_t = q$ and is the following:

$$sh_W^p = \frac{s^p(\frac{1-\alpha}{\beta})}{s\left(r+\frac{1-\alpha}{\beta}\right)-s^p.r^p} sh_{Y_L}^p \quad (1)$$

The only change comes from the wealth-income ratio β and the share of capital income α that become respectively $\beta = \frac{W}{Y} = \frac{s(1+q)}{g-q}$ and $\alpha = \frac{rW}{Y} = r.\beta = \frac{r.s(1+q)}{g-q}$ at the steady state.[45]

Starting from the steady-state formula (1) and using the new expression of β and α, we end up with:

$$sh_W^p = (1+\frac{s^p r^p - sr}{\frac{g-q}{1+q} - s^p r^p})\frac{s^p}{s} sh_{Y_L}^p \quad (5)$$

As compared to the steady-state formula (2), the introduction of an exogenous rate of capital or depreciation mitigates the impact of growth on wealth concentration.

[45] At the steady-state, $\frac{W_{t+1}}{Y_{t+1}} = \frac{W_t}{Y_t}$. Using $W_{t+1} = (1+q)[W_t + sY_t]$ and $Y_{t+1} = (1+g)Y_t$, we obtain $\beta = \frac{s(1+q)}{g-q}$ and $\alpha = \frac{rW}{Y} = r.\beta = \frac{r.s(1+q)}{g-q}$.

Appendix F. Reconciliation using household wealth surveys

Appendix F deals with the reconciliation of wealth surveys with aggregate national accounts and our benchmark series based on the capitalization method. The Excel File GGP2016WealthAppendixF includes the main appendix tables and figures relative to this reconciliation. All the tables and figures are derived from Stata codes included into the folder DataFiles (see ReadMe file in directory). We present in this section the French wealth surveys along with the methodology used to adjust them.

Section F.1. Description of wealth surveys

The objective of the wealth surveys is to describe the household situation with regard to financial, real-estate and professional assets and liabilities in France. It provides also a description of socio-demographical characteristics of the households as well as household income, gifts and inheritances received during lifetime.

The French institute of statistics and economic studies (Insee) compiled wealth surveys every six years starting in 1986 (1986, 1992, 1998, 2004, 2010 and 2014).[46] Since 2010, the French wealth survey is part of the eurosystem household finance and consumption survey (HFCS) that harmonizes the wealth surveys of the 15 euro area countries. From 2014, the French wealth survey will be computed every three years and be partially panelized. As of July 2016, the final files of the 2014 wealth survey is not available yet.

Several aspects of the survey have been improved over time. The first wealth surveys (1986 and 1992) include the value of household wealth by wealth brackets rather than exact amounts. In the wealth surveys 1998 and 2004, the simulated residual method was used to estimate exact amount from the answers by wealth brackets.[47] Since 2010, households are asked to evaluate the exact amount of their

[46] These wealth surveys were called « enquête actifs financiers » in 1986 and 1992, and « enquête patrimoine» since 1998.

[47] In the wealth survey 1998, the level of the different assets and the level of gross wealth were estimated separately. Consequently, the sum of the different assets is not consistent with the total gross wealth.

assets.[48] The 2010 wealth survey is of relatively high quality due to several improvements. It is matched with income tax returns and benefits from a large oversampling at the top based on taxable wealth information.[49] However, its sample size is still too small to go beyond the 99th percentile.

Section F.2. Reconciliation with national accounts and capitalization method

Tables F1.a, F1.b and F2 from the Excel file GGP2016AppendixF presents summary statistics of wealth inequality, wealth composition by wealth groups and the wealth-survey reporting rates - i.e. ratios (wealth survey)/(national accounts) - by asset category for year 1992, 1998, 2004 and 2010.[50] At this stage, we did not try to adjust the wealth distribution of the wealth surveys. The only adjustments we made consisted in individualizing the data and match it with the total adult population.[51]

Table F1.a shows clearly the impact of the better oversampling on the estimation of top wealth shares in the wealth survey 2010. Top 0.1% wealth share ranges from 3% to 4% between 1992 and 2004 and increases dramatically to 8% in 2010 due to the better oversampling procedure.

Total wealth reported in the wealth surveys corresponds to 60%-70% of aggregate private wealth from national accounts. Non-financial assets (housing and business assets) are typically well covered by the surveys with a reported rate of 80% to 90%. In contrast, only 30%-40% of total financial assets are reported in the wealth surveys.

---

[48] When individuals refuse or are unable to evaluate precisely their wealth, they are asked to range their wealth by wealth brackets and the simulated residuals method is used to estimate the exact amount.

[49] Due to these major changes, top wealth shares obtained from the 2010 wealth survey is much more reliable than those from the previous waves.

[50] We did not use the wealth survey of 1986 for comparability issues. In this survey, there is no distinction between housing and business assets and the financial debts are out of the scope of the survey.

[51] We keep only individual adults (20 years old or more) and split equally the net wealth among married couples. The adult population covered by the wealth surveys corresponds usually to 95%-98% of the total adult population from the Census Data. We corrected for this discrepancy by multiplying each individual weight by the ratio of the total adult population in the Census Data over that of the wealth survey. This very simple adjustment has the advantage to be entirely neutral in terms of wealth inequality. In the wealth survey 1992, wealth and its components are reported by wealth bracket. We use a uniform law for the intermediate brackets and a Pareto law with a coefficient of 2 for the last wealth bracket to recover the exact amount.

This discrepancy can be due to misreporting or non-response at the top. A recent literature has tried to better estimate top wealth shares in wealth surveys using national accounts, billionaire rankings and Pareto interpolation techniques (see in particular Vermeulen (2014, 2016) and other references provided in Blanchet 2016).

We propose in this section an alternative methodology to correct top wealth share from wealth surveys using our benchmark series based on the capitalization method. This exercise is therefore closely related to Saez and Zucman (2016) and Bricker et al. (2016) that try to reconcile administrative and wealth survey data.

We first reconcile wealth surveys with national accounts in Table F3 assuming that the only source of discrepancy between the sources is due to uniform misreporting within asset class. This is done by simply dividing each asset reported at the individual level by the wealth-survey reporting rates - i.e. ratios (wealth survey)/(national accounts) - corresponding to the asset category.[52] Note that by definition this adjustment let unchanged the asset-specific distributions. As equities and bonds are much less reported in the wealth surveys that housing assets (Table F1.b) and the share of financial assets increases with the level of wealth (Table F2), the adjustment increases the wealth concentration. Consequently, top 1% wealth share increases from 14%-16% to 16%-21% after the adjustment during the 1992-2004 period.

Figure F1 shows that the uniform adjustment within asset class closes almost entirely the gap between top 10% wealth shares derived from the wealth surveys and our benchmark series. While the gap for top 1% wealth is almost entirely closed for years 1992 and 2010, it is only divided by two for years 1998 and 2004 (Figure F2). The remaining discrepancy can be explained by non-response at the top. Uniform adjustment within asset class performs poorly when there is a differential non-response. This is particularly of concern when non-response is more pronounced at the top, whose wealth composition differs strongly from the rest of the distribution. Indeed, the big gap of the top 1% wealth share between 2004 and 2010 due to a

[52] We define 7 asset classes in the wealth surveys (gross housing assets, debt, business assets, equities and investment fund shares, bonds and loans, deposits and savings accounts, life insurance and pension funds) that are perfectly consistent with their counterpart in the national accounts.

better oversampling of the top wealth holders in the wealth survey 2010 at the top remains after the uniform adjustment.

In order to correct for both misreporting and non-response at the top, we adopt a more sophisticated adjustment method varying by asset and asset-specific distribution. Tables F4 and F5 show the asset-specific distributions over time based on the wealth surveys and the capitalization method. These tables show that the asset-specific distributions of equities and bonds differ strongly. In particular, equities and bonds are much more concentrated at the very top in the capitalization method as compared to the wealth surveys. These assets are indeed mainly concentrated at the top of the wealth distribution, which is affected by non-response in the wealth surveys. In contrast, the other asset-specific distributions (Owner-occupied housing assets, life insurance assets, household debt, and deposits and savings accounts) are very similar. They are indeed less concentrated at the top and therefore less affected by non-response at the top.

We then estimate in Table F6 the wealth-survey reporting rates - i.e. ratios (wealth survey)/(national accounts) - by asset category and by group of asset-specific distribution. We use these ratios to reconcile the wealth surveys with our benchmark series. As it turns out, we only need to estimate differential reporting rate for equities and bonds. The results are presented in Table F7. Figures 1 and 2 show that the adjusted top 10% and top 1% wealth shares (using the differential adjustment) are consistent both in trend and level with those from the capitalization method. Note that the reconciliation is partly obtained by construction: owner-occupied housing, life insurance assets, household debt, and deposits and savings accounts are imputed in the capitalization method (W1) using wealth surveys (see imputation methods in Appendix B).

Appendix G. Reconciliation using annual wealth tax data

Appendix G deals with the reconciliation of wealth tax data with aggregate national accounts and our benchmark series based on the capitalization method. The Excel File GGP2016WealthAppendixG includes the main appendix tables and figures relative to this reconciliation. All the tables and figures are derived from Stata codes included into the folder DataFiles (see ReadMe file in directory). We present in this section the tabulations from the French wealth tax along with the methodology used to adjust them.

Section F.1. Description of wealth tax data

The French government introduced a progressive annual tax on top wealth holders (approximately the top 1%[53]) in 1982 (Impôt sur les Grandes Fortunes (IGF)). The wealth tax was abolished in 1986, reintroduced in 1989 (Impôt de Solidarité sur la Fortune (ISF), and is still in place in 2016. Table G1 reports the evolution of the wealth tax schedule (thresholds and marginal tax rates by tax brackets) from 1982 to 2015.

For this study, we have collected wealth tax tabulations by tax brackets from 1984 to 2013. Although micro-files of wealth tax data are not publicly available, wealth tax tabulations are computed on an irregular basis by the French tax administration for various parliamentary reports or other official public reports related to taxation.[54] Tables G2 and G3 present the number of individuals and the average wealth by tax brackets since 1984 as reported in the official public reports. In these tables, we report as well the thresholds of each tax bracket and the inverted Pareto coefficients both by tax brackets and for the cumulative distribution above the threshold.

These data suffer from important limitations due to the characteristics of the French wealth tax.

[53] At its creation in 1982, only the top 0.5% wealth holders were affected by the wealth tax. As wealth has increased much more rapidly than the tax threshold, the proportion of wealth holders subject to the wealth tax increased continuously from 0.5 % in 1982 to 2.1 % in 2010.

[54] The official public reports are generally produced either by the French parliament (Senat or Assemblee Nationale) or by independent public administrations such as Conseil d’Analyse Economique or Conseil des Prélèvements Obligatoires (formerly called Conseil des Impôts).

Tax evasion:

The French wealth tax is a self-assessed tax: households have to evaluate whether they are subject to the wealth tax, estimate the market value of their net assets by themselves (there is no third-party reporting), compute the corresponding taxes, and send the tax return to the tax administration. These features are likely to create three types of tax evasion. First, households with wealth just above the tax threshold eligibility have strong incentives to underreport the value of their assets to avoid wealth taxation as the probability of audits is low for this group.[55] Second, there is no legal definition of market value. People can therefore underestimate, intentionally or not, the market value of their assets in periods of strong capital appreciation. Third, households can simply ignore their eligibility to the wealth tax.

Tax-exempt assets

French wealth tax is characterized by many full or partial tax-exemptions. As the wealth tax tabulations report only the taxable wealth, the exempted fraction of the assets is not reported in the tabulations.

The first source of tax exemption is related to the broad category of professional assets. All assets necessary for the carrying-out of a professional activity are entirely exempted from wealth tax and therefore not reported in the wealth tax returns. The professional assets include also equity participations in family firms and in companies where assets holders play an active management role under certain conditions.

The second source of tax exemption is related to housing assets. The primary residence of the households benefits from a tax exemption of 30 %.[56] The tenant-occupied housing can benefit from a tax exemption of 20 % if unfurnished. Tenant-

[55] Audits are based on an algorithm that checks the consistency of the wealth tax returns. It is based on the value of the assets reported during the past three years. New eligible households to wealth tax are therefore less likely to be targeted by audits. Moreover, the algorithm does not assess the potential underreporting of housing assets, which are the main component of the wealth owned by new eligible households to the wealth tax.

[56] A tax exemption of 20 % for the primary residence was introduced in 1996. This exemption went up to 30% in 2008.

occupied housing benefits from a fully tax exemption, when the households rent furnished apartments as a consequence of their professional activity.

Finally, specific life insurance assets can also benefit from partial or full tax exemption.[57]

Section F.2. Adjustment and reconciliation of wealth tax data

Correction for tax evasion in the first tax bracket

As described in the previous section, important tax evasion behaviors around the tax eligibility may arise from the self-assessed feature of the wealth tax. Indeed, Table G2 shows that the inversed Pareto coefficient is 25% higher in the first tax bracket than in the next three ones, denoting that there are less taxpayers around the tax threshold eligibility than they should be. In addition, the threshold of the top 1% of the wealth distribution corresponded to the threshold of the first tax bracket in 2000 and to the threshold of the second tax bracket in 2007. As the Pareto coefficients are relatively stable over time, the Pareto coefficient of the second tax bracket in 2007 should have been relatively similar to that of the first tax bracket in 2000. In contrast, only the Pareto coefficient of the first tax bracket remains significantly higher than the next one, while the tax bracket eligibility corresponds to a different moment of the distribution (P98 in 2007 instead of P99 in 2000).

We correct for tax evasion in the first tax bracket by simply assuming that the "true" Pareto coefficient of the first tax bracket is equal to that of the second tax bracket. Using the features of the Pareto law, we are able to estimate the corrected average wealth and the corrected number of individuals in the first tax bracket. Tables G4 and G5 show the harmonized wealth tax tabulations after correction for tax evasion in the first tax bracket using this approach (see also Zucman, 2008).

---

[57] Only contributions made to non-redeemable policies such as "contrat d'assurance-vie à bonus de fidélité" or "contrat de capitalisation" have to be reported in the wealth tax returns. Indeed, capital income and capital gains generated by these specific life insurance policies are not considered as part of the household wealth. Term insurance policies are entirely tax exempted from the wealth tax.

Estimation of the unadjusted wealth shares from wealth tax tabulations: 1984-2013

Table G6 reports the evolution of the top 1% wealth share from wealth tax tabulations during 1984-2013 period at the individual and at the household levels.

To do that, we simulate the top 3% of the wealth distribution at the household level using the Pareto law and its parameters by tax brackets such as reported in Tables G2 and G3. More precisely, a Pareto distribution has a cumulative distribution function of the form:

$$F(w) = 1 - \left(\frac{w_{min}}{w}\right)^{a}$$

Where w is the level of wealth considered, F(y) represents the fraction of individual with wealth above w and w_{min} is the minimum level of wealth (the threshold of P97 or the different tax brackets) over which the distribution follows a Pareto law of coefficient a.

For each year over the period 1984-2013, we draw a distribution of 200,000 observations with equal weight such as the total distribution represents the top 3% wealth holders. We then estimate the level of wealth of each observation from different tax brackets using the following equation:

$$w = \frac{w_{min}}{1 - F(w)^{1/a}}$$

In order to individualize the distribution of top wealth holders, we randomly draw households that can be either married couples or single individuals such that 80% of households are married couples. We then replace each household corresponding to married couples by two individuals and divide their wealth by two.

Reconciliation of the wealth tax data with our benchmark series

Figure G1 depicts the evolution of the top 0.1%, top 0.5-0.1% and top 1-0.5% wealth shares over the period 1984-2013 derived from the wealth tax tabulations. In Figures G2 to G5, we compare all these wealth shares with those from our benchmark scenario derived from the capitalization method. These figures show a big difference between the series. Our benchmark top 1% wealth shares are 2 to 3 times higher than those obtained using wealth tax tabulations. The large differences between the series can be explained by the large exemptions to the wealth tax that reduce dramatically the level of wealth reported in the wealth tax returns (see above for the detailed description of the tax exemption to the wealth tax).

By making plausible assumptions on the fraction of tax-exempt wealth by asset categories and levels, we are able to reconcile this data with our benchmark estimates. Tables G8 and G9 report the unadjusted and the corrected wealth shares as well as the fraction of tax-exempt wealth by asset categories that we apply in order to correct the raw series directly derived from the wealth tax tabulations. These correcting factors vary by wealth groups (top 0.1%, top 0.5-0.1% and top 1-0.5%). More specifically, we first estimate an exempted fraction of each asset corresponding to the exemption rules of the wealth tax, to possible under-estimation of the asset and to tax evasion.[58] For each wealth groups, we estimate the fraction of wealth that is exempted by multiplying each asset component by the corresponding coefficient. We then add the exempted part of wealth to the taxable one in order to get the corrected wealth share.

Figures G2 to G5 compares the unadjusted wealth shares with the corrected ones and with our benchmark wealth shares derived from the capitalization method for the top 1%, top 0.1%, top 0.5-0.1% and top 1-0.5%. The figures show that the level and the evolution of the corrected wealth shares are very close to those of our benchmark series. But there is significant uncertainty about the exact level and evolution of tax exemptions, so it is difficult to use this source on its own. In particular, we stress that the reconciliation that we propose here is reasonably plausible, but is by no means the only possible reconciliation.

[58] For the top 1% wealth share, we take the following tax exemption coefficients : 100% for business assets, 30% for owner-occupied housing assets since 2008 (20% between 1996 and 2007 and 0% before), 20% for tenant-occupied housing assets since 1996 (0% before), an additional 10% exemption for housing assets (tenant and owner-occupied) due to possible tax evasion and under-reporting, 70% for equities and debt assets (this fraction corresponds to assets that are used for professional activities) and 70% for life insurance. See Tables G8 and G9 for more details on the tax-exempted fraction used by wealth groups and asset categories.

Appendix H. Reconciliation using the French rich families list (*Challenges* list)

The Appendix H relates to the reconciliation of the rich family ranking with benchmark series based on the capitalization method. The Excel File GGP2016WealthAppendixH includes the main appendix tables and figures relative to this reconciliation. All the tables and figures are derived from Stata codes included into the folder DataFiles (see ReadMe file in directory). We present in this section the French rich families list along with the methodology used to correct for its limitations.

Section H.1. Description of the "*Challenges*" list

In 1983, Forbes started to publish a list of the 400 richest Americans (the "*Forbes 400*"), followed in 1987 by the *World's Billionaires List*, a list of 140 billionaires whose 96 were out of the US. Since then, this list of the world's wealthiest people has been published annually. Following this example, newspapers in other countries have started to publish billionaire lists. In France, the weekly magazine "*Challenges*" began to establish a list of France's 500 wealthiest families in 1996.

Methodology of the Challenges list

As explained by one of the journalist in charge of the French list[59], the methodology relies on different steps. A first step is to establish the list of wealthiest people. Journalists first look at the financial reports to find the wealthiest French stock holders. This kind of information is generally public and easily available. Unfortunately, most of the richest families' wealth relies on unquoted shares (just one third of the families referenced in the list have their wealth composed with quoted shares). Therefore, journalists have to go through professional publications, seminars, awards ceremonies and all kind of events where successful and richest businesspersons meet. In 2012, their database contained more than 3,000 names. Once the family names are gathered, the last step is to evaluate their wealth. For quoted shares, the number of shares is simply multiplied by its market value. For

[59] All the following information rely on an article of the Challenges magazine giving some methodological precisions (in French): http://www.challenges.fr/entreprise/comment-evalue-t-on-leur-patrimoine_4006.

unquoted shares, journalists examine turnover, balance sheets, etc., of the company in order to compare it to publicly traded companies or to recent transactions within the same business sector. Finally, journalists send letters to the rich people who are likely to be part of the list. The mail includes an estimation of their wealth and asks them to correct, confirm or precise some points. According to the *Challenges* journalists, some of these rich people cooperate, although we do not know very precisely how many do so.

Drawbacks of the Challenges list

This list provides valuable additional information on very top holders but presents some drawbacks that have to be accounted for when using it to compare with other series. First, the methodology previously described, though "well-oiled" and carefully applied, does not guarantee an accurate measurement of wealth. Second, the concept of wealth used in the *Challenges* list is not net wealth but gross wealth. This is not necessarily the most important drawback since at this level of wealth it should be close to net wealth. Third, this list is a ranking of families and does not provide the number of members of each family. It is thus not possible to compare it directly with our benchmark individual estimates. Fourth, some of the families included in this list do not live in France and should therefore not be included in our French wealth series. This can lead to a serious overestimation of the French top 0.001% wealth share.

Section H.2. Reconciliation with capitalization method

Tables H1 from the Excel file GGP2016AppendixH presents the top 0.001% wealth share derived either from capitalization method or *Challenges* lists from 1998 to 2013. As previously explained, some limitations of the *Challenges* list have to be accounted for. We insist on the fact that our own corrections also suffer from serious limitations, due to the difficulty to find accurate statistical information both on the production and on the bias of this list. Our corrections have to be seen as first steps towards a better understanding of the adjustments needed to lead to comparable results.

First, in order to individualize the family wealth reported in the *Challenges* list, we gather information about family size in the top wealth distribution thanks to the Forbes *25 America's Richest Families List*. We thus rely on information for the US since, to our knowledge, no such information can be found for the French richest families. It turns out that the median family size is 11 and the mean 14. We alternatively consider that each family include 1, 10 or 15 members that we add in our sample and divide the family wealth equally among all the members. Then, we have to take into account that each member of the family can be married. In order to compute individual wealth on a similar basis to our benchmark estimates, we randomly draw members of the families that can be either married couples or single individuals such that 80% of these individuals are married.[60] We then replace each family members that are married by two individuals and divide their wealth by two.[61]

Second, we use external information about non-French resident families in order to correct the *Challenges* list. We mainly rely on articles from economic and financial newspapers. As it turns out, 100 families in the top 500 list are not French residents and live in Belgium, Luxembourg, UK or Switzerland.[62] We roughly evaluate that non-resident wealth represents around 20% of the total wealth of the top 500 list.[63] Once again, this is clearly a rough estimate. We consider it as a lower bound of non-resident wealth and further work should be done to provide a better estimate. Therefore, to not overestimate the top wealth share in France, we correct for this non-French resident wealth by subtracting it to the total wealth reported in the *Challenges* list.

[60] This figure corresponds to the share of individuals in couple such as estimated from our top 0.001% wealth share series.
[61] Basically, we first correct for family size (dividing wealth by 10 or 15 and creating individuals holding this wealth). Then we randomly draw 80% of these individuals for whom we re-divide wealth by 2 and create a spouse holding the same amount of wealth.
[62]http://www.challenges.fr/economie/20130709.CHA1936/exil-fiscal-grosse-vague-de-depart-dans-le-top-500-des-plus-grandes-fortunes-de-france.html
[63] 49 families are described as living in Switzerland (holding 55 billion euros, see e.g http://www.leparisien.fr/economie/exil-fiscal-49-francais-parmi-les-300-plus-riches-de-suisse-28-11-2014-4330035.php#xtref=https%3A%2F%2F) and 20 in Belgium (see e.g. http://www.lecho.be/actualite/economie_politique_belgique/Les_Francais_detiennent_17_milliards_d_euros_en_Belgique.9544615-3158.art?ckc=1). Since we did not find any other information (about families living in the UK or Luxembourg for instance), we just take into account those two elements to compute the share of non-resident wealth.

Last, once split into individuals, the 500 richest families do not represent a constant share of the population across time. This clearly makes comparison difficult. To keep comparable top wealth shares, we select for each year as many individuals as needed to represent 0.001% of total population.[64]

Figure H1 shows the effect of considering different family sizes on the top 0.001% wealth share. Our higher correction for family size (15 individuals) leads to a top 0.001% wealth share still a bit higher than our benchmark estimate with capitalization method. It is normal since, at this stage, non-French resident wealth has not been excluded. Figure H2 presents our preferred comparisons, after correction for non-French residents. As it turns out, our benchmark estimates give consistent levels of wealth share when compared to our preferred corrected *Challenges* series. Again, in the same way as with the wealth tax data, we stress that these reconciliation assumptions are reasonably plausible, but are certainly not the only possible ones. Our general conclusion is that these sources – wealth taxes and billionaire rankings – offer interesting complements to other sources, but suffer from too many uncertainties to be used on their own.

[64] The top 0.001% represented 500 individual in 2010.

References

ALVAREDO, F., ATKINSON, A., CHANCEL, L., PIKETTY, T., SAEZ E. and ZUCMAN, G. (2016), "Distributional National Accounts (DINA): Concepts and Methods used in WID.world", *WID.world Working Paper* (http://www.wid.world), version V1.

ATKINSON, A. and HARRISON, A.J. (1978), *Distribution of Personal Wealth in Britain, 1923-1972 (Cambridge University Press)*.

BLANCHET, T. (2016), “Wealth inequality in Europe and in the United States: estimations from surveys, national accounts and wealth rankings”, *PSE Working Paper*.

BLANCHET, T., FOURNIER, J. and PIKETTY, T. (2017), “Generalized Pareto curves: Theory and applications”, *PSE Working Paper*.

BLANPAIN, N. (2016), “Les inégalités sociales face à la mort – Tables de mortalité par catégorie sociale et diplôme”, Insee Résultats, série Société, n° 177.

BOURDIEU, J., POSTEL-VINAY, G. and SUWA-EISENMANN, A. (2003) "Pourquoi la richesse ne s'est-elle pas diffusée avec la croissance ? Le degré zéro de l'inégalité et son évolution en France : 1800-1940", *Histoire et mesure*, **23(1-2)**, 147-198.

BOURDIEU, J., KESZTENBAUM, B. and POSTEL-VINAY, G. (2013), *L’enquête TRA, histoire d’un outil, outil pour l’histoire* (*Tome I : 1793-1902*, INED Editions).

BRICKER, J., HENRIQUES, A.M, KRIMMEL, J.A and SABELHAUS, J.E (2016), “Measuring Income and Wealth at the Top Using Administrative and Survey Data”, *Brookings Papers on Economic Activity*, **47(1)**, 261-331.

BUISSON, G. and LAPINTE, A. (2013), “Le couple dans tous ses états”, Insee Première, n° 1435.

EUROSTAT, IMF, OECD, UN, WB (2009), *System of National Accounts 2008*, New York.

GARBINTI, B., GOUPILLE-LEBRET, J. and PIKETTY, T. (2017), "Income Inequality in France, 1900-2014: Evidence from Distributional National Accounts (DINA), *PSE Working Paper*.

HFCS (2015), Household finance and consumption survey. Statistical micro-level database, https://www.ecb.europa.eu/home/html/researcher_hfcn.en.html, Accessed June 2 2015, European Central Bank.

LANDAIS, C., PIKETTY, T. and SAEZ, E. (2011), *Pour une révolution fiscale - Un impôt sur le revenu pour le 21e siècle* (Seuil).

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

PIKETTY, T. (2010), "On the Long-Run Evolution of Inheritance: France 1820-2050", Data Appendix, *PSE Working Paper*.

PIKETTY, T. (2011), "On the Long-Run Evolution of Inheritance: France 1820-2050", *Quarterly Journal of Economics*, **126(3)**, 1071-1131.

PIKETTY, T. (2014), *Capital in the 21st century* (Harvard University Press).

PIKETTY, T., POSTEL-VINAY, G. and ROSENTHAL, J.L. (2004), "Wealth concentration in a developing economy: Paris and France, 1807-1994", *CEPR Discussion Paper*.

PIKETTY, T., POSTEL-VINAY, G. and ROSENTHAL, J.L. (2006), "Wealth concentration in a developing economy: Paris and France, 1807-1994", *American Economic Review*, **96(1)**, 236-256.

PIKETTY, T., POSTEL-VINAY, G. and ROSENTHAL, J.L. (2014), “Inherited vs self-made wealth: theory and evidence from a rentier society (Paris 1872-1927)”, *Explorations in Economic History*, **51(1)**, 21-40.

SAEZ, E and ZUCMAN, G. (2106), “Wealth Inequality in the United States since 1913: Evidence from Capitalized Income Tax Data”, *Quarterly Journal of Economics*, **131(2)**, 519-578.

SCF (2015), Survey of consumer finances, Accessed December 2 2015, Federal Reserve, http://www.federalreserve.gov/econresdata/scf/scfindex.html

VERMEULEN, P. (2014), “How Fat is the Top Tail of the Wealth Distribution?”, *European Central Bank Working Paper*.

VERMEULEN, P. (2016), "Estimating the Top Tail of the Wealth Distribution", American Economic Review: Papers & Proceedings, **106(5)**, 646-50.

ZUCMAN, G. (2008), “Les hauts patrimoines fuient-ils l'ISF?”, *PSE Working Paper*.