

Wage Inequality and Labor Market Outcomes in France

Long-run Trends and Comparison with Wealth Inequality and other Countries

Alix Myczkowski

Supervisor:

Thomas Piketty

Referee:

Facundo Alvaredo

Paris School of Economics
Master Analysis and Policy in Economics

August 2018

Abstract

This study aims at examining the distribution of labor income in France from 1913 to 2014. It relies on the analysis of fiscal data, matched with surveys for most recent years, using generalized Pareto interpolation techniques.

It appears that the distribution of wages remained very stable over the 20th century in France, especially within the bottom and the middle of the distribution. This confirms that the decline of income inequality in France is largely imputable to the evolutions of capital income inequality. The rise of top wages in the private sector since the years 2000 could mark the start of a new trend of rising labor income inequality. For the moment, the distribution of labor income remains much more unequal among self-employed workers than among wage earners.

JEL Classification: N34

Keywords: Economic History, Wages, Labor Markets, Income Inequality

Acknowledgements

I would first like to thank my supervisor Thomas Piketty, for his insightful comments and useful guidance throughout the pursuit of this work. I would also like to thank Facundo Alvaredo for accepting to be referee. I am also very grateful to Thomas Blanchet for his helpful advice, and to Clémentine Van Effenterre and Stéphane Myczkowski for their constructive comments.

Contents

List of Figures	1
Section 1. Introduction	2
Section 2. Related Literature	5
Section 3. Data sources and Methods	7
Section 4. Long-run trends and short-term fluctuations of labor income inequality in France	16
Section 5. Directions for Future research	61
Section 6. Conclusion	65
REFERENCES	67
APPENDIXES	69

List of Figures

Graph 1 - Average Wages - Main Groups (1919 - 2015).....	16
Graph 2 - Average Wages in France - Detailed Groups (1919 - 2015).....	17
Graph 3 - Wage Shares - Main Groups (1919 - 2015)	19
Graph 4 - Wage Shares - Detailed Groups (1947 - 2015)	20
Graph 5 - Percentiles - Main Groups (1919 - 2015).....	21
Graph 6 - Percentiles - Detailed Groups (1919 - 2015).....	22
Graph 7 - Wage Shares for Civil Servants - Main Groups (1913 - 1996).....	24
Graph 8 - Wage Shares for Civil Servants - Detailed Groups (1913 - 1996).....	25
Graph 9 - Wage Shares - Main Groups (1919 - 1919)	27
Graph 10 - Percentiles - Main Groups (1913 - 1919).....	28
Graph 11 - Wage Shares - Main Groups (1919 - 1938)	29
Graph 12 - Percentiles - Main Groups (1919 - 1938).....	30
Graph 13 - Wage Shares within the Top 1% for Civil Servants (1919 - 1996).....	33
Graph 14 - Percentiles for Civil Servants (1913 - 1996).....	35
Graph 15 - Wage Shares within the Top 10% (1947 - 2015).....	37
Graph 16 - Pareto Coefficients (1919 - 2015)	38
Graph 17 - Labor Income Shares for Labor Income Earners (1996 - 2014)	44
Graph 18 - Top Labor Income Shares for Labor Income Earners (1996 - 2014).....	45
Graph 19 - Top Percentiles for Labor Income Earners (1996 - 2014)	46
Graph 20 - Labor Income Shares for Private Sector Wage Earners (1996 - 2014).....	47
Graph 21 - Top Labor Income Shares for Private Sector Wage Earners (1996 - 2014)	48
Graph 22 - Labor Income Shares for Public Sector Wage Earners - Main Groups	49
Graph 23 - Labor Income Shares for Public Sector Wage Earners - Detailed Groups (1996 - 2014)	50
Graph 24 - Labor Income Shares for Self-employed Workers (1996 - 2014).....	51
Graph 25 - Top Labor Income Shares for Self-employed Workers (1996 - 2014)	52
Graph 26 - Top Percentiles for Self-employed Workers (1996 - 2014).....	53
Graph 27 - Labor Income Shares for Retirees and Unemployed (1996 - 2014)	54
Graph 28 - Labor Income Shares in Net Wages and Pre-tax Labor Income (1970 - 2015) ...	57

Section 1. Introduction

With the explosion of very high wages in the Anglo-Saxon world over the last decades, especially in the United States of America, there has been growing interest for labor income inequalities in the economic literature and the public debate.

In the recent literature about income inequality (Piketty 2014, among others), it has been shown that labor income inequalities are generally less stringent and more stable than capital income inequalities. Furthermore, the bulk of the decrease in income inequality in most developed countries over the 20th century is imputable to the decline in capital income inequalities, because capital stocks were severely affected by the two successive world wars and by following economic and fiscal policies. This contradicts an earlier theory according to which economic growth “naturally” triggers a decrease of income inequalities, in particular through sectorial transfers of the labor force (Kuznets 1955). Thus, the stability of labor income inequalities is a central issue in the economic theory of inequalities.

Though they are less important than capital income inequalities, labor income inequalities are significant in the developed and the developing world, and their recent explosion in Anglo-Saxon countries has revealed their potential as a driver of income inequalities. Such a phenomenon hasn't taken place in France, but there has been a significant rise of top remunerations there as well. More generally, with the collapse of great fortunes over the 20th century, the share of labor income has become much more important at the top and the very top of the income distribution, in France and elsewhere. Thus, the potential for labor income to drive a new long-run trend of increasing income inequalities is also a crucial issue in the literature and a current concern in the public debate.

The literature on wages in France (Baudelot et Lebeauvin 1979, Bayet 1997, Bayet et Julhès 1996, among others) paid limited attention to dispersion indicators until the years 2000, as it focused mostly on categorical indicators, such as the average wage of various categories of workers (factory workers, employees, qualified workers etc.). In his work on the evolution of top incomes in France, T. Piketty (2001) used employer tax returns to follow the evolution of top wages and of their relation to the average wage from 1919 to 1996. Piketty, Garbinti and Goupille (2017), also used fiscal summary statistics and combined them with detailed income tax micro-files, available for the years 1970 and after, to estimate the distribution of labor income among the entire adult French population from 1970 to 2014.

Both works conclude that despite non-negligible short-run fluctuations, the distribution of labor income has remained very stable over the 20th century.

This study aims at examining in a systematic manner the entire distribution of labor income and its evolution over the 20th century. I rely on employer tax returns as well, and use the generalized Pareto interpolation techniques developed by Blanchet, Fournier and Piketty (2017) to estimate the distributions. I use the fact that summary statistics on private wages cover all workers since 1947 to estimate precisely the entire distribution of wages as early as 1947, and extrapolate it for the preceding years. I also use complementary data on civil servants to compare wages in the private and the public sector. For the recent period (1996 – 2014), the *Enquête Revenus Fiscaux et Sociaux* (ERFS), that consists of individual tax returns matched with the *Enquête Emploi* (a nationally representative survey on work-related issues), enables to decompose the distribution of labor income between four groups: private sector wage earners, public sector wage earners, self-employed workers, and retirees and unemployed, in order to compare labor income inequalities across sectors and work statuses.

The results of this study confirm the remarkable stability of the distribution of labor income in France over the 20th century. In particular, the repartition of wages was even more stable within the bottom 90% of the distribution, and most short-run fluctuations of wage inequalities were imputable to evolutions in the top 10%, or even the top 1%. Only civil servants experienced a major reduction of wage dispersion in the middle of the century. If their appears to be a non-negligible increase in the income share owned by the top 1% of the distribution over the last decade, there is no explosion of very high wages in the private sector, as there has been in the United States since the years 1970. As for the recent period, our decomposition shows that wages are less spread out than self-employed labor income, and less spread out in the public than in the sector.

The results for the recent period, in particular the closeness between the distribution of labor income for all sectors and work statuses and the distribution of wages in the private sector, are encouraging regarding the possibility of using the long-run series on wages constructed in this work to build long-run series on the distribution of labor income among the entire adult population.

The rest of this paper is organized as follows. Section 2 presents the literature on labor income in France. Section 3 describes the data sources and estimation methods used in this

work. Section 4 analyzes the results. Section 5 presents ideas for future research. Section 6 concludes.

Section 2. Related Literature

The evolution of the distribution of labor income in France over the 20th century is an issue at the intersection of two literatures: the literature on wages and labor income in France, and the literature on income inequality.

The literature on the evolution of wages in France (Baudelot and Lebeauvin 1979, Bayet 1997, Bayet and Julhès 1996, among others) resorts to inequality indicators, such as deciles ratios. But it mainly focuses on the comparison of the average wage of different categories of workers (especially between the average wage of factory workers and the general average), and on identifying the impact of the structural changes of the labor force on these wages. In terms of methods, most papers rely on employer tax returns.

These works agree on the following periodization: after a moderate progression during the first half of the century, wages really took off between 1950 and 1976, and have been progressing much more moderately since then. As to the dispersion of wages, the periodization is also well-agreed on: wage dispersion increased significantly between 1950 and 1967, reduced between 1968 and 1983, and has slightly increased again (but mostly stagnated), since then. Overall, the literature agrees on the stability of the wage distribution.

According to the pioneer work of Kuznets (1955), income inequality should decrease following economic growth, as workers move towards more productive sectors. This theory has been challenged by a new literature on income inequality since the beginning of the years 2000 (Piketty 2001, 2003, 2014, Piketty, Garbinti and Goupille 2017, regarding France, among others). This literature argues that income inequalities do not follow any structural trend, and that their decline in various countries, such as France, over the 20th century is mostly imputable to the decrease in capital income inequalities. The two world wars resulted in important destructions of physical capital, and fortunes affected by these destructions could not be rebuilt because of the economic and fiscal policies of the second half of the 20th century.

The study of labor income inequality is thus central, as a necessary corollary of this thesis is that the distribution of labor income remained relatively stable over the 20th century, which has indeed been the case in France according to Piketty (2001, 2014) and Piketty, Garbinti and Goupille (2017).

The same literature also underlined the rising potential of labor income inequalities to drive a new long-run increasing trend of income inequalities. Piketty (2014) shows that in a vast number of countries the share of labor income at the top of the distribution has grown significantly (though capital income is still the main source of income at the very top). Furthermore, a number of studies (Piketty and Saez 2003 for instance) have thrown light on the explosion of labor income inequalities in Anglo-Saxon countries, especially the United States, since the years 1970, driven by the rise of very high wages disconnected from any productivity rationale. It appears that France, despite a substantial increase of very high remunerations in recent years, hasn't experienced a similar widening of the wage dispersion. But the question remains whether Anglo-Saxon countries are an isolated case or whether other countries, such as France, are on the same path a few decades late.

The recent literature on labor income inequalities in France (Piketty 2001, 2003, 2014) has on the one hand thrown light on the evolution of top wages over the 20th century, and their relation to the average wage, using summary statistics from employer tax returns. On the other hand, Piketty, Garbinti and Goupille (2017) have constructed series on the distribution of pre-tax labor income for the entire adult French population from 1970 to 1914, using income tax micro-files. All these studies conclude that the distribution of labor income in France remained very stable. However, the possibility of estimating precisely the entire distribution of wages since 1947, using data from employer tax returns, has not been exploited yet, and neither has the very rich information on workers from the *Enquête Emploi* matched with individual tax returns for recent years.

Section 3. Data sources and Methods

Concepts

This study deals with three concepts of labor income: net wage, fiscal labor income, and pre-tax labor income.

Net wage corresponds to the amount received by employees from their employers (with taxable indemnities), net of pension contributions, contributions to unemployment insurance, and contributions to the social security system. It doesn't include in kind benefits and non-taxable indemnities. It is not equivalent to "taxable" wage, because it corresponds to the wage before any tax deduction has been applied. The main long-run series used in this study are expressed in net wage.

Fiscal labor income corresponds to the labor income reported on their tax returns by income earners, net of pension contributions and other social contributions. The main difference with net wage is that it includes other types of labor income: it is the sum of wages, unemployment benefits, retirement pensions, and of 70% of mixed income. Mixed income is income that has both a capital and a labor component: it corresponds to self-employed workers' income (industrial and commercial profits, non-commercial profits, and agricultural profits¹, respectively called BIC, BNC and BAG in the rest of this paper), ancillary revenues ("revenus accessoires"), "revenus imposés au quotient" (income taxed in a specific way), and income perceived abroad but taxed in France ("revenus perçus à l'étranger"). I chose to discount mixed income by 0.7 following the assumption adopted in Piketty, Garbinti and Goupille (2017), that the labor component of mixed income amounts to 70% of total income. Thus fiscal labor income concerns a wider range of workers than net wages. The data source used for the recent period (1996 – 2015) enables to build series expressed in fiscal labor income.

The concept of pre-tax labor income is part of the larger framework of "Distributional National Accounts", forged by the WID.world team. It corresponds to "the sum of all income

¹ In French, respectively « Bénéfices Industriels et Commerciaux » (BIC), « Bénéfices Non

flows going to labor (...), after taking into account the operation of the pension system, but before taking into account other taxes and transfers” (Piketty, Garbinti and Goupille 2017). Its main difference with fiscal labor income is that the non-deductible *Contribution Sociale Généralisée* (CSG²) is added at the individual level, and that production taxes accruing to labor income, a proportion of undistributed profits and corporate income taxes accruing to the government, and a proportion of the net primary surplus and net capital income of government and NPISH³ are added. These proportions are added under the assumption that these national accounts aggregates fall proportionally on the different income categories, the details of the procedure are explained by Piketty, Garbinti and Goupille (2016, 2017). The results of this work are compared with pre-tax labor income series constructed by Piketty, Garbinti and Goupille (2017) for the adult French population from 1970 to 2014.

Data sources

Data sources for long-run net wage series (1919 – 2015)

The long-run net wage series presented in this work all rely on the same kind of source: wage tax returns filled by employers, who had to declare to the fiscal administration the annual amount of net wages paid to each of their employees.

Wage tax returns appeared first with the implementation of the schedular wage tax (*impôt cédulaire sur les traitements et salaires*) in 1917. The French tax administration started compiling summary statistics of these declarations at the national level in 1919. It did so for each year until 1938, because in 1939 the schedular tax was transformed into a tax deducted at source. However, the primary source, employer tax returns, continued to exist, and took the name of *Déclarations Annuelles de Données Sociales* (DADS). The French national statistical institution, the *Institut National de la Statistique et des Études Économiques* (INSEE), started compiling again national summary statistics from these declarations in 1947, and produced them for almost every year between 1950 and 1999 (INSEE tables are called DADS tables in the rest of the paper). Since 2000, INSEE doesn't publicly release similar tables anymore, but

² The *Contribution Sociale Généralisée* is a tax created in 1991, deducted at source.

³ Non-profit institutions serving households

publishes distributional tables derived from the same source instead (at the decile level until 2010 and at the centile level since 2011).

Interwar national tables (constructed by the French tax administration for years 1919 to 1938) and DADS tables (constructed by INSEE for years 1947 to 1999) have very similar contents. They are both expressed in net wage, except from 1931 to 1938, where tables are expressed in taxable wage (wage to which tax deductions have been applied). They are organized the same way: for a large number of wage brackets, are indicated the total number of workers whose wage falls into that bracket, and sometimes the total amount of wages paid to all these workers.

There are a few differences between the two periods however. First, the population covered is not the same. Interwar tables cover only taxable workers (the first threshold of the table isn't 0 but the schedular wage tax threshold), but from any sector (public and private). So these tables capture only the top of the distribution of all wage earners. On the other hand, DADS tables cover the entire wage distribution (the first threshold of the tables is always 0), but for part of the private sector only: they exclude domestic, agricultural and part-time workers for most of the period, as well as non-permanent workers from 1953 to 1962. Another important difference is that DADS tables only report headcounts and bracket thresholds, (except for 1947, 1951, 1952, and from 1993 to 1999). So they display no information on the total of wages paid in each bracket, and thus no information on the average wage per bracket.

This kind of data has various assets. First, it can be deemed very reliable. Employers deduced these wages from the profit tax base they declared to the tax administration, so they didn't have any incentive to minimize them. On the other hand, the information on employer returns were used to check workers' individual declarations, for whom it was essential that the wages were not over-declared. The balance between these two opposite incentives implies that the amounts reported are probably very close to reality.

Second, it enables to study the entire wage distribution as early as from 1947, whereas statistics compiled from general income tax returns remained restricted to taxable households for the whole 20th century. Thus, this data source is a valuable tool to study the evolution of income in France in the second half of the 20th century.

This data has several limits however. In terms of temporal coverage, there are no tables between 1939 and 1946, and the tables publicly released by INSEE from 2000 to 2010 are not detailed enough to yield reliable estimations, so we are not able to present results for these years.

Moreover, before 1947, the coverage is very limited (from 5% of wage-earners in 1919 to 37% in 1938). This problem is reinforced by the existence of various tax deductions for married individuals and families with children or dependents from 1922 to 1933. Individuals whose wage was above the tax threshold but who were not taxable thanks to tax deductions were not counted in interwar tables. This issue concerns a significant number of individuals, as, for instance, a married man with two children would benefit from a deduction of 9000 francs in 1927, which would bring his tax threshold to 16 000 francs instead of 7000 francs. Besides, interwar tables only included full-time workers, and thus couldn't capture the impact of partial unemployment, an important consequence of the economic crisis of the years 1930.

Even after 1947, DADS tables probably overestimate wages at the bottom of the distribution, because they exclude the workers whose wages were potentially the lowest (domestic, agricultural and part-time workers). Domestic and agricultural workers disappeared over the second half of the 20th century, but the number of part-time workers has been increasing significantly since the years 1980. Moreover, civil servants are excluded. More generally, interwar and DADS tables don't include other types of workers than wage earners (in particular self-employed workers), nor individuals who get an income thanks to past work (unemployed and retirees).

In terms of content, the absence of any bracket level information in DADS tables makes the estimation of the wage distribution less precise. Besides, top thresholds are not always high enough to cover the very top of the distribution (the top 0.1% and sometimes the top 1%) properly.

More details on these sources can be found in Appendixes A and B.

The only correction I made concerned the tables from 1922 to 1933, where individuals are missing because of tax deductions for family situations. The detailed explanation of the correction procedure is in Appendix A, I only develop the general principle here.

There were three types of tax deduction: one for married couples, one per children, one per adult dependent. For highest brackets, it can be considered that no one is missing (wages are too high for people to become non-taxable thanks to deductions). Thus I used the number of tax deductions per type of deduction in these brackets to compute headcounts ratios between

various family categories (the ratio between the number of married individuals with no child and the number of married individuals with one to three children, and so on).

For most brackets, some categories of families are missing but not all, because the deduction associated to certain categories are too low to enable them to be not taxable: only numerous families are missing in upper-middle brackets for instance. Assuming constant ratios between family categories⁴ over brackets, I apply the ratios computed with the highest brackets to the “full” categories (where no one is missing) of lower brackets.

For low brackets, where there is no “full” family category left, I assumed linear repartition of wages within each bracket to impute missing individuals. (That is to say: if married individuals without children were not taxable if their earnings were under 13 000 francs, I divided the number of married individuals without children of the (10 000 – 20 000) bracket by 0.7). The lowest brackets could sometimes not be corrected: when no one from the category with the lowest deduction (married individuals without children) is counted in the table, because all individuals from that category and that bracket are not taxable, linear imputation cannot even be applied. Consequently, these brackets had to be dropped.

From 1931 to 1933, no systematic correction could be made, because the tables are expressed in taxable wage (so someone whose fiscal wage was in a certain bracket could be in an inferior bracket in the table), and there was no information permitting to come back to fiscal wage and impute missing individuals. Thus, following a method used by Piketty (2001), I only upgraded post-estimation results in an ad-hoc manner (detailed in Appendix A), to account for lower taxable wages and so as to not disturb long-run trends.

These main data sources were complemented with similar data on public workers for four years: 1913, 1931, 1951, and 1996. This data comes from various sources (detailed in Appendix D), such as budget reports for the earliest years, and civil servant censuses and surveys for the more recent years.

They have the exact same format as DADS tables: they indicate for a large number of earnings brackets the number of public workers whose wage falls into that bracket, and are expressed in net wage. These tables cover the whole distribution, and ensure a good coverage of both the top and the bottom of the distribution. However, they do not display any bracket-level information.

⁴ This assumption was already made by Piketty (2001) to correct for missing people in the general income tax summary statistics.

Data sources for the recent period (1996 – 2015)

From 1996 to 2015, the *Enquête Revenus Fiscaux et Sociaux* (ERFS) enables to decompose the distribution of labor income between various subgroups.

The ERFS results from the match of the *Enquête Emploi* with individuals' tax returns. The *Enquête Emploi* is a nationally representative survey conducted by INSEE, on work-related issues (types of jobs, unemployment, working conditions, remunerations, etc.). It was annual until 2002 and is quarterly since then. The base sample of ERFS is made of all the individuals surveyed in the *Enquête Emploi* at a given year whose household could be associated to at least one tax return. Consequently, not all the observations of ERFS have filled a tax return, but at least one member of their household has. All those whose household could not be associated with any tax return were dropped (but the potential specificity of dropped observations is corrected for). Thus, ERFS benefits from the strengths of survey and fiscal data: income data is very reliable and precise (it is displayed at the household, fiscal unit, or individual level, depending on the variable), and the information on jobs (work status, sector, etc.) is very rich.

Income data enables to construct a labor income variable expressed in fiscal labor income, as defined above. I also added non-deductible CSG that is usually not included in fiscal income because it is not taxable, because it enables the variable to be closer (in terms of concept) to pre-tax national income.

I only kept individuals over 20 years old, and used variables from the *Enquête Emploi* to decompose labor income earners in four groups: wage earners from the public sector, the private sector, self-employed workers, retirees and unemployed. This decomposition was straightforward as there are precise questions on work status⁵ and sector. Public wage earners include State and local community workers only, whether they are civil servants, contract workers or interns, but not workers from state services such as transportation companies or social security institutions (what is sometimes called the semi-public sector). Private wage earners include all others wage earners, including temporary agency workers (*travailleurs intérimaires*), apprentices, interns, non-permanent workers, and also company executives if they earn a wage (*salariés chef d'entreprise*). Self-employed workers include individual

⁵ Individuals are asked whether they are employees, self-employed, retirees, unemployed etc.

workers, employers who do not earn a wage, and family helpers (*aide familiaux*). Retirees and unemployed also include former self-employed workers who have a pension (*anciens retirés des affaires*). A benchmark group is also created, made of all adults who earn a labor income superior a third of the nominal minimum wage (computed each year).

I chose to keep the same labor income concept for all subgroups. For instance, labor income for self-employed workers doesn't correspond to the sum of their BIC, BNC and BAG, but to the concept explained above⁶, so the distribution of labor income for self-employed isn't strictly equivalent to the distributions of BIC, BNC and BAG. Thus if a wage earner also gets profits from a free-lance activity (consulting for instance), these are counted. This solution doesn't enable to display the differences between the distributions of the various types of labor income. On the other hand, it matches reality better, as it shows what different categories of workers actually receive: if a wage earner also has an income from a free-lance consulting activity, it means its paid job lets him enough time and energy to have this activity.

For each worker category and for the benchmark group, I built tables similar to the interwar tables (for a large number of income brackets, the tables display the number of workers and the total wage bill of the bracket), in order to estimate separate distributions per sectors. The main difference is that the brackets are much more numerous than in DADS and interwar tables (and cover well the top of the distribution), so they enable more precise estimation. I also built separate distributions by income concept: for private wages, public wages, BIC-BNC-BAG, and pensions, to compare with the by-worker-type distributions. The results are overall quite close.

Generalized Pareto Interpolation

To estimate full distributions of wages and labor income, I applied to each of the aforementioned tabulations the generalized, non-parametric Pareto interpolation techniques developed by Blanchet, Fournier and Piketty (2017), using for this the “gpinter” computer codes developed by the same authors and available online.

The Pareto distribution is the following:

⁶ Sum of wages, pensions, and 70% of mixed income

$$1 - F(y) = \left(\frac{k}{y}\right)^a$$

F is the repartition function (that is to say the probability for income x to be inferior to y). The main property of the Pareto distribution is that the ratio $y^*(y)/y$, with $y^*(y)$ the average income of all observations above income level y, is equal to a constant $b = a/(a-1)$, called the inverted Pareto coefficient, independent of the income level y. The Pareto distribution has often been used to model the distribution of income in the literature on income repartition (Pareto 1897, Kuznets 1953). However, assuming a constant parameter b can permit to model properly the top of the distribution, but is too simplistic for the entire distribution.

The generalized Pareto interpolation method estimates the whole distribution of income by interpolating the full curve of inverted Pareto coefficients $b(p)$, (p is the percentile), from tables such as the ones in our data. There is no parametric constraint, allowing $b(p)$ to vary over the distribution to match closely the input data.

When no bracket level information on average is reported (which is the case from 1956 to 1992 and for civil servants), the results rely on the interpolation of the quantile function, and are bound to be less precise than with information on both thresholds and averages. When no bracket level information on thresholds is provided, the results rely on the interpolation of the Lorenz curve, and are also less precise.

The full description of the techniques can be found in the article of Blanchet, Fournier and Piketty (2017).

Post-interpolation corrections

Two important corrections were made on post-interpolation results (percentiles, earnings levels and earnings shares).

First, for the years 1919 to 1938, to provide approximate estimates of what happened within the bottom 90% of the distribution, I assumed that the bottom 50% and the middle 40% of the distribution each had the same share of the bottom 90%⁷ than in 1947 (the first

⁷ That is to say the ratios : (Middle 40% share/Bottom 90% share) and (Bottom 50% share/Bottom 90% share) are constant from 1919 to 1947.

year when this share could be properly estimated). Thus, their share of the total wage bill varied proportionally to the share of the bottom 90% (which can be properly estimated for the whole interwar period). This assumption, called “proportional split”, enables to derive income shares for the bottom 50% and the middle 40% of the distribution for the interwar period.

I then use the share-based generalized Pareto interpolation technique to obtain results on percentiles below the 90th percentile for the years 1919 to 1938. For the top 10%, I kept the results obtained from the first interpolation, which was more precise as it was a share-and-thresholds interpolation.

These hypothetical bottom shares and percentiles enable to understand what would have happened if the repartition of labor income within the bottom 90% had remained constant from 1919 to 1947. This assumption is certainly too simplistic, but it appears that, at least for the following years of the 20th century, the repartition of wages within the bottom 90% actually remained relatively stable, so it is probably not too far from reality.

Secondly, I corrected post interpolation results for the years 1953 to 1962, to correct for the exclusion of non-permanent workers in DADS tables.

For 1952 and 1963, there are tables for permanent workers only (group A) and tables for all workers (group B), so I could estimate the wage distribution for both groups. For each percentile and average income of the subgroups of the distribution studied, I derived the ratio of their value for group B on their value for group A, in 1952 and 1963. Assuming that these ratios evolved linearly (whether they decreased or increased) between 1952 and 1963 value, I interpolated equivalent ratios for each year between 1952 and 1963. I then applied these ratios to the percentiles and income levels I had for permanent workers only, to get estimations for all workers. The resulting income shares do not sum up to 1, because ratios are computed independently for each group of the distribution, I adapt the top 10% share to other shares to have consistent shares that add up to 1. That is to say that the top 10% share is equal to 1 minus the sum of all other shares. It probably results in underestimating the top 10% share, this is discussed in Appendix B.

Finally, I corrected for missing years in the DADS data (1948, 1949, 1953, 1955, 1958, 1981, 1983 and 1990) by linear interpolation of percentiles and average wage levels, (the correction procedure for 1954 is explained in Appendix B).

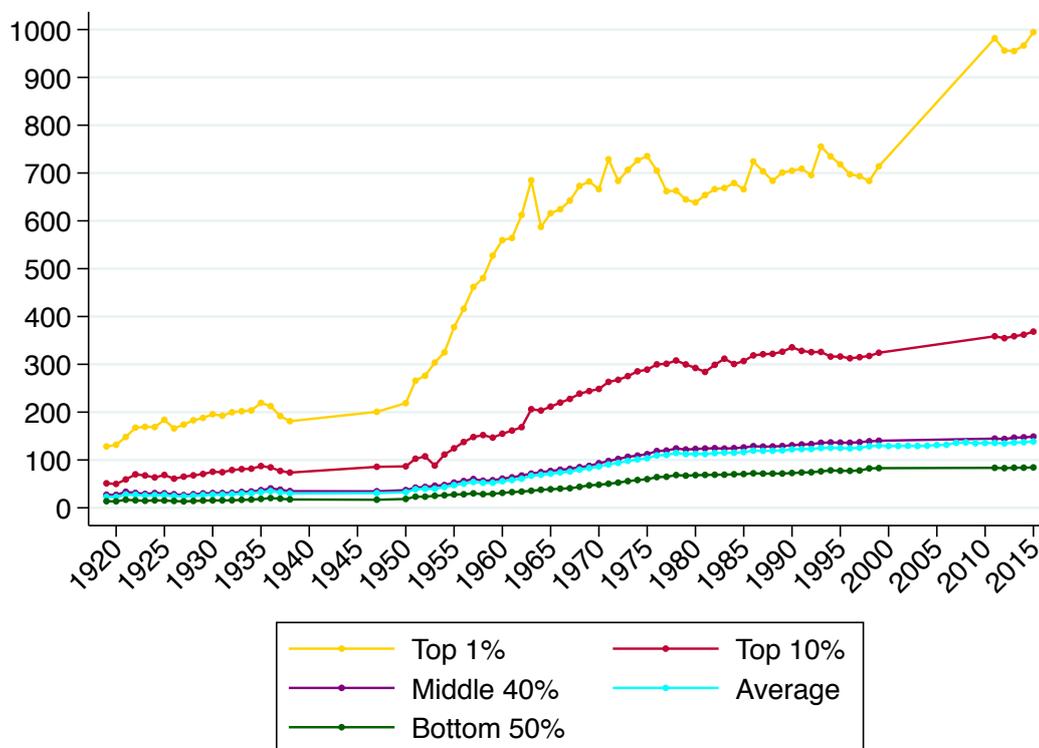
Section 4. Long-run trends and short-term fluctuations of labor income inequality in France

Long-run trends

Average wages

Graph 1 - Average Wages - Main Groups (1919 - 2015)

(in thousands of francs 1998)



Notes: Graph 1 displays the average wage of subgroups of the wage distribution:

- The Top 1% gathers individuals whose wage is above the 99th percentile of the distribution.
- The Top 10% gathers individuals whose wage is above the 90th percentile of the distribution.
- The Middle 40% gathers individuals whose wage is between the 50th and the 90th percentiles of the distribution.
- The Bottom 50% gathers individuals whose wage is inferior to the 50th percentile of the distribution.
- The Average is the national average wage.

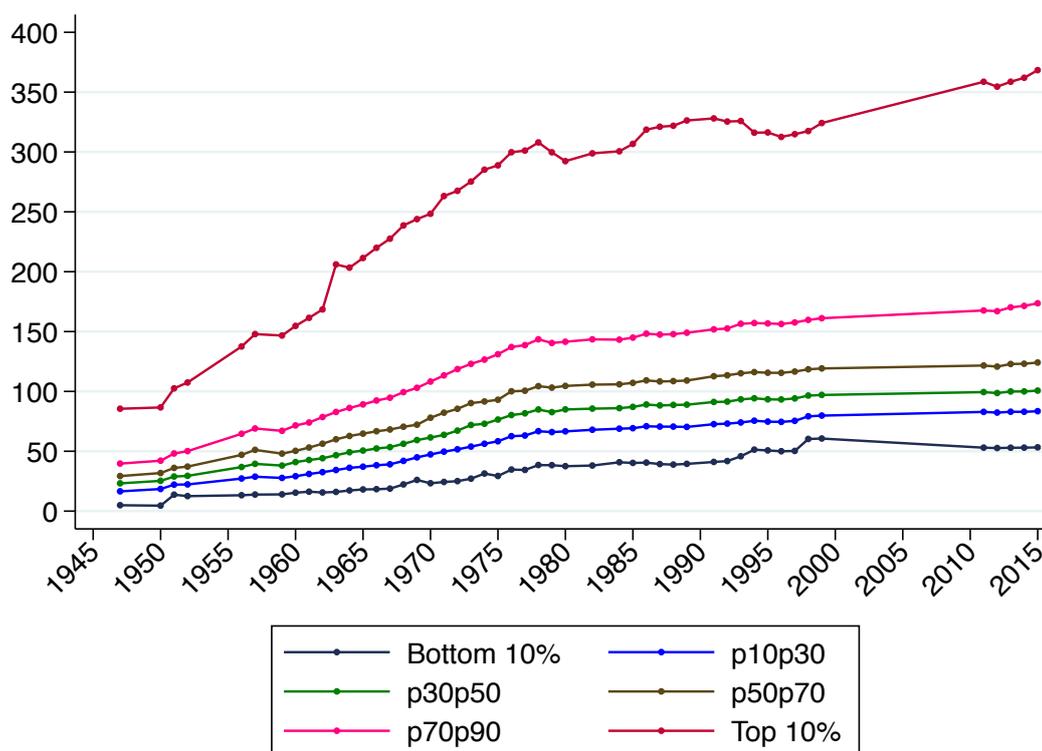
Wages are expressed in thousands of constant francs 1998. All sectors are covered from 1919 to 1938, only the private sector is covered between 1947 and 2015.

Sources: Tax administration summary statistics on the Impôt céduaire sur les salaires (1919 – 1938), Déclarations Annuelles de Données Sociales (DADS) (1947 – 2015).

Meaning: In 1965, the average wage of the top 10% of the distribution is equal to 211 390 francs of 1998.

Graph 2 - Average Wages in France - Detailed Groups (1919 - 2015)

(in thousands of francs 1998)



Notes: Graph 2 displays the average wage of subgroups of the wage distribution:

- The Bottom 10% gathers individuals whose wage is inferior to the 10th percentile.
- p10p30 gathers those whose wage is between the 10th and the 30th percentiles of the distribution.
- p30p50 gathers those whose wage is between the 30th and the 50th percentiles of the distribution.
- p50p70 gathers those whose wage is between the 50th and the 70th percentiles of the distribution.
- p70p90 gathers those whose wage is between the 70th and the 90th percentiles of the distribution.
- The Top 10% gathers those whose wage is above the 90th percentile of the distribution.

Wages are expressed in thousands of constant francs 1998. All sectors are covered from 1919 to 1938, only the private sector is covered between 1947 and 2015.

Sources: Tax administration summary statistics on the *Impôt cédulaire sur les salaires* (1919 – 1938), DADS (1947 – 2015).

From what can be seen on Graphs 1 and 2, the average wage of the whole population stagnated or increased moderately from 22 820 to 34 755 constant francs between 1919 and 1936, decreased slightly between 1936 and 1947, and increased almost constantly, at a high rate, from 30 750 constant francs in 1947 to 114 723 in 1978. From 1978 onwards, it continued to increase but at a much smaller rate (it was 138 510 constant francs in 2015).

With a few (non-negligible) variations, the average wages of the groups of the distribution have followed the same path, with the same turning points. This periodization is in line with the existing literature: the high growth rate from 1950 to 1970 is linked to the exceptional economic growth France has experienced during the period known as the “Trente Glorieuses”, which stopped in the years 1970 and has been much slower since then.

It can be noted however, that the gap between the average wages of the top 10% and 1% of the distribution and other subgroups really widened between 1950 and the middle of the years 1970, as the average wages of the top 10% and top 1% increased much more (in absolute value) than for the rest of the distribution. Thus, when their growth slowed down in 1978, like it did for the rest of the distribution, this gap was much wider and it remained so. The same remark can be made for the gap between the wages of top 10% and the top 1%. The average wage of the top 1% is also the only one to experience another “take-off” in the years 2000, not followed by the top 10%. Besides, the average wages of both these groups seem to be more volatile (even more for the top 1%): only they experience significant decreases in value over short periods of times (like at the end of the years 1970).

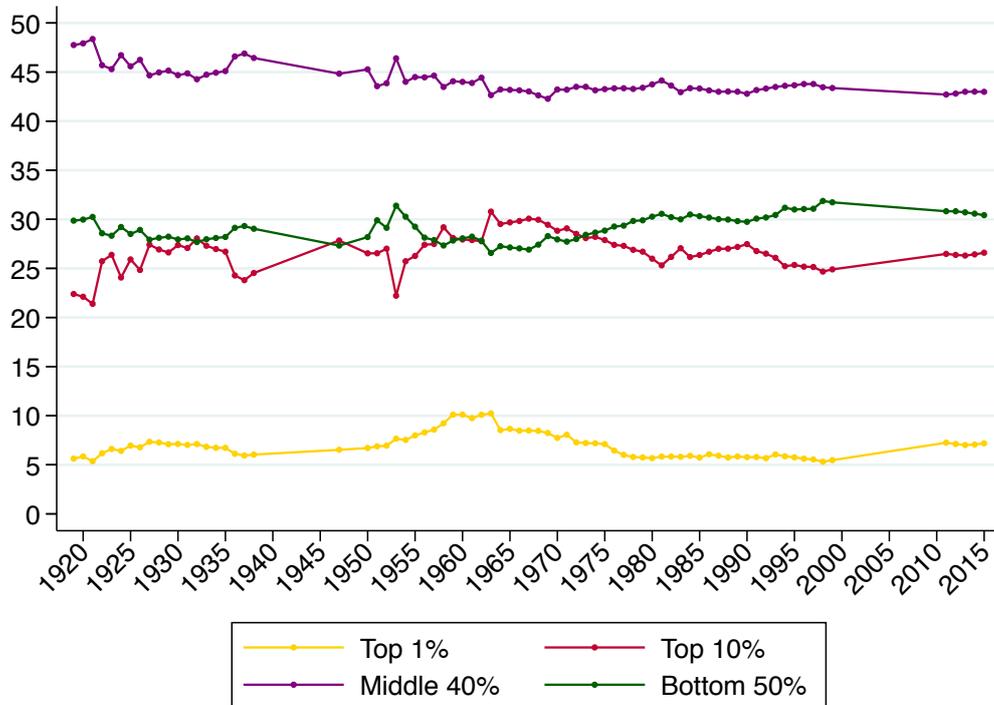
This is all the more remarkable as the whole bottom 90% of the distribution seems to evolving in a similar way, even when looking at smaller subgroups on Graph 2: there seems to be a real difference between the top 10% and the bottom 90% in terms of “behavior”. Within the bottom 90%, it can be noted that the bottom 10% started increasing significantly only in the middle of the years 1960, increased more slowly (in value) than the rest, except in the years 1990, and even decreased during the years 2000.

However, the evolution in value of subgroups’ average wages doesn’t say anything about the evolution of inequalities, which is why we now turn to wage shares, that is to say the proportion of the total national wage bill earned by each subgroup of the distribution.

Income Shares

Graph 3 - Wage Shares - Main Groups (1919 - 2015)

(in percentage of total wage bill)



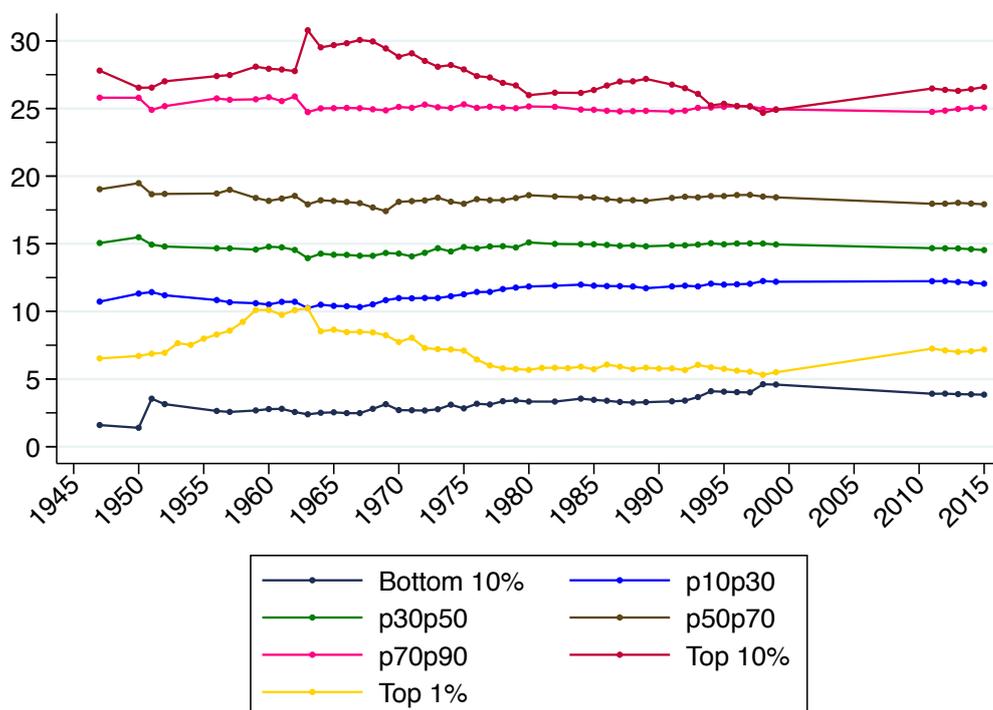
Notes: Graph 3 displays the ratios of the total of wages paid in each subgroup of the distribution on the total national wage bill. See Graph 1 for the definitions of subgroups. All sectors are covered from 1919 to 1938, only the private sector is between 1947 and 2015.

Sources: Tax administration summary statistics on the *Impôt cédulaire sur les salaires* (1919 – 1938), DADS (1947 – 2015).

Meaning: In 1980, about 30% of all the wages paid in France (in the private sector) went to individuals from the Bottom 50% of the distribution.

Graph 4 - Wage Shares - Detailed Groups (1947 - 2015)

(in percentage of total wage bill)



Notes: Graph 4 displays the ratios of the total of wages paid in each subgroup of the distribution on the total wage bill, in the private sector. See Graph 2 for the definitions of subgroups.

Sources: DADS

The first striking conclusion we can draw from Graph 3 is the long-run stability of income shares, for all the subgroups of the distribution. Indeed, the income share of the top 10% ranges between 21% and 32% over the whole period, these boundaries are 5% and 10% for the top 1%, 42% and 46% for the middle 40%, and 26% and 32% for the Bottom 50% (and the intervals are even narrower by dividing the middle 40% and the bottom 50% in smaller groups).

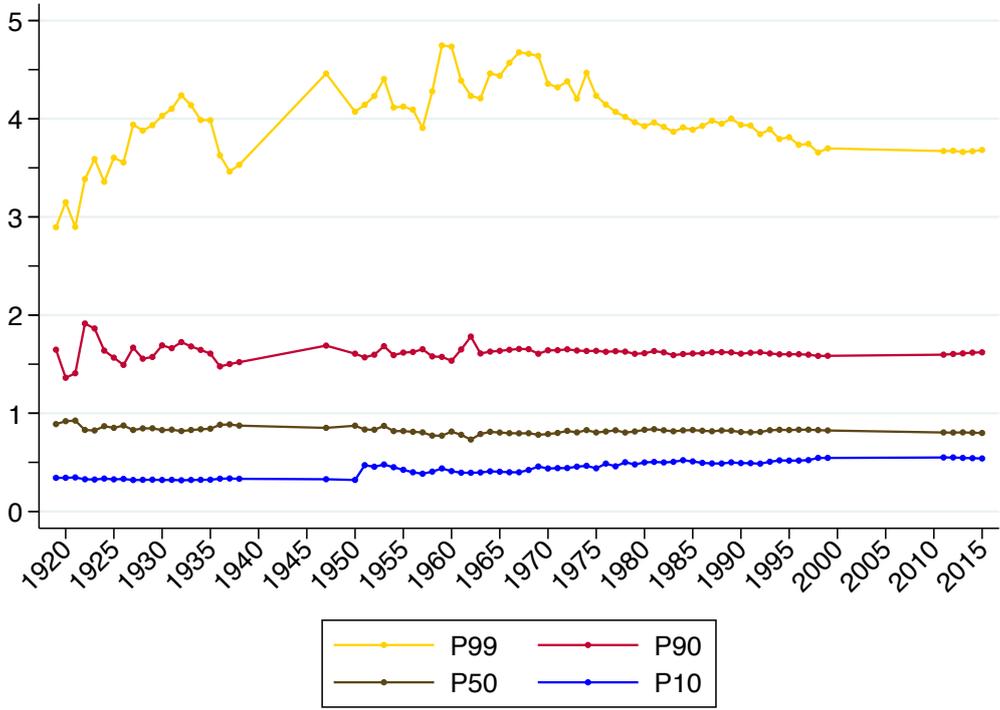
Fluctuations between these various boundaries are not negligible at all, especially when they occur over 10 or 20 years. When the share of the top 10% increases from 26% in 1950 to 30% in 1967, it means that the 10% richest wage earners used to earn 2.6 times the average wage in 1950 and earn about 3 times the average wage 17 years later, which is a significant transformation, especially when it occurs over an individual career.

However, the main teaching of Graphs 3 and 4 is that there is no apparent long-run trends in labor income inequality: even if the top 10% wage share increased up to more than 30% in the years 1960, at the end of the century it is back to its average level of the years 1920 (it is equal to 24.9% both in 1926 and 1999). The same can be said for the top 1% wage share: it

climbs to about 10% at the beginning of the years 1960, but it is equal to 5,5% in 1999 like in 1919 (5,6%). Only the share of the bottom 10% appears to have gradually increased from 1.6% in 1947 to 4.5% in 1999, but this long-run rise might have come to an end as it appear to have decreased slightly since 2000 (3.8% in 2015). It can again be noted that there is a clear demarcation between the top 10% (and the top 1%) and the rest of the distribution in terms of income share volatility: the “consequences” of the top 10% share’s fluctuations seem to be rather equally distributed among the subgroups of the bottom 90%, so that each subgroup is much less volatile.

The stability of wage inequality is confirmed when looking on Graphs 5 and 6 at percentiles (expressed in proportion of the average wage), which are usually more volatile than shares. Except P99 and P10 (respectively the minimum income levels needed to be in the top 1% and the top 90% of the distribution), all thresholds look remarkably stable (P90 is volatile during the interwar but it doesn’t follow any trend).

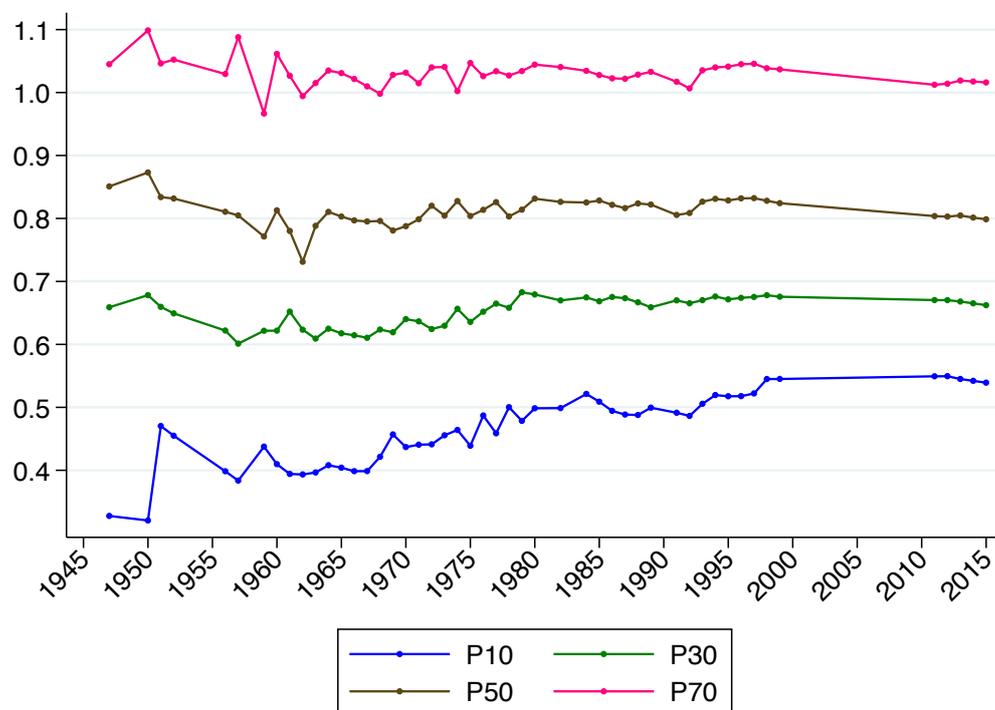
Graph 5 - Percentiles - Main Groups (1919 - 2015)
(in proportion of average wage)



Notes: Graph 5 displays percentiles of the wage distribution, expressed in proportion to the national average wage. P99, P90, P50 and P10 are respectively the 99th, 90th, 50th, and 10th percentiles. All sectors are covered from 1919 to 1938, only the private sector is between 1947 and 2015.
Sources: Tax administration summary statistics on the *Impôt cédulaire sur les salaires* (1919 – 1938), DADS (1947 – 2015).
Meaning: In 1937, the 90th percentile is approximately equal to 1.5 times the average wage.

Graph 6 - Percentiles - Detailed Groups (1919 - 2015)

(in proportion of average wage)



Notes: Graph 6 displays percentiles of the wage distribution, expressed in proportion to the national average wage, in the private sector. P30 and P70 are respectively the 30th and 70th percentiles.
Sources: DADS (1947 – 2015).

The stability of the French labor income distribution is all the more remarkable as labor markets have been deeply transformed over the period: wage labor became the main form of work, and two major sectorial transfers of the labor force occurred (from agriculture to industry and from industry to services). The general level of qualifications of the workforce increased unprecedentedly during the second half of the century, and women massively entered the labor market in the years 1980. Despite these crucial evolutions, the income shares have only fluctuated within 5 or 10 percentage point intervals.

There are several explanations for this. One of the main explanatory theories for wage inequalities is the relation between the supply and the demand of qualifications: when a qualification is highly demanded but scarcely supplied, its price (which corresponds to the worker's wage) increases. In France, no dramatic demographic evolution disturbed the equilibrium between the supply and demand of qualifications, and even if the supply

increased, all qualifications levels increased at a similar rhythm, while the demand for qualifications increased in parallel. Thus, in the end the relation between the supply and the demand of qualifications was not modified, and despite the increase in average wage for all subgroups of the distribution, inequalities remained constant.

Tolerance towards high and very high wages provides another possible explanation. According to Piketty (2003), various signs indicate that inequalities within the labor income distribution have always been relatively overlooked, compared to inequalities between owners and non-owners of capital. Indeed, top wage earners have long been associated to the “classe moyenne” (middle class), as if they were not at the top of the income distribution. Even public policies seem to have neglected wage inequalities: the general income tax has most of the time targeted people at the very top of the distribution (that is to say capital owners) rather than people at the “lower top” (who correspond to very top wage earners). As a consequence, the idea according to which some wages are “too high” is very recent whereas very high wages are not new, and there has not been much attempt to control them, and thus reduce the range of the distribution.

The stability of the wage distribution has several implications. First, it goes against Kuznets’ theory (1955) according to which economic growth necessarily triggers in the long run a decrease in economic inequalities. It also confirms that the secular decrease of income inequality over the last century in France is mostly imputable to the decrease in capital income inequalities (Piketty 2001, 2014, Piketty, Garbinti and Goupille 2017).

In terms of inequality levels, it can be seen on Graph 3 that the 10% richest wage earners’ share in the total wage bill has always been approximately equal to the share of the poorest half of wage earners. The 1% richest paid workers have always earned on average 5 to 10 times the national average wage. This means that France is not one of the most equal countries (in Scandinavian countries, the top 10% share fluctuates around 20%), but not one of the most unequal either (in the United States, the top 10% share was about 35% in 2010⁸).

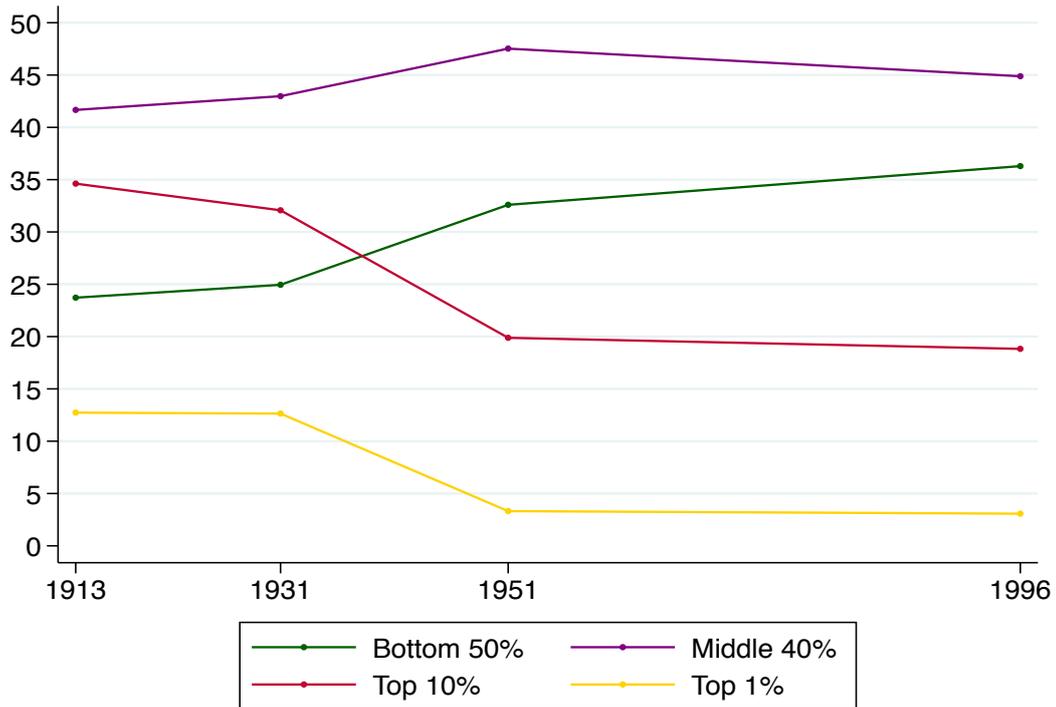
⁸ Figures from Piketty (2014)

Long-run trends for civil servants

Regarding civil servants, long-run trends are strikingly different.

Graph 7 - Wage Shares for Civil Servants - Main Groups (1913 - 1996)

(in percentage of total wage bill)

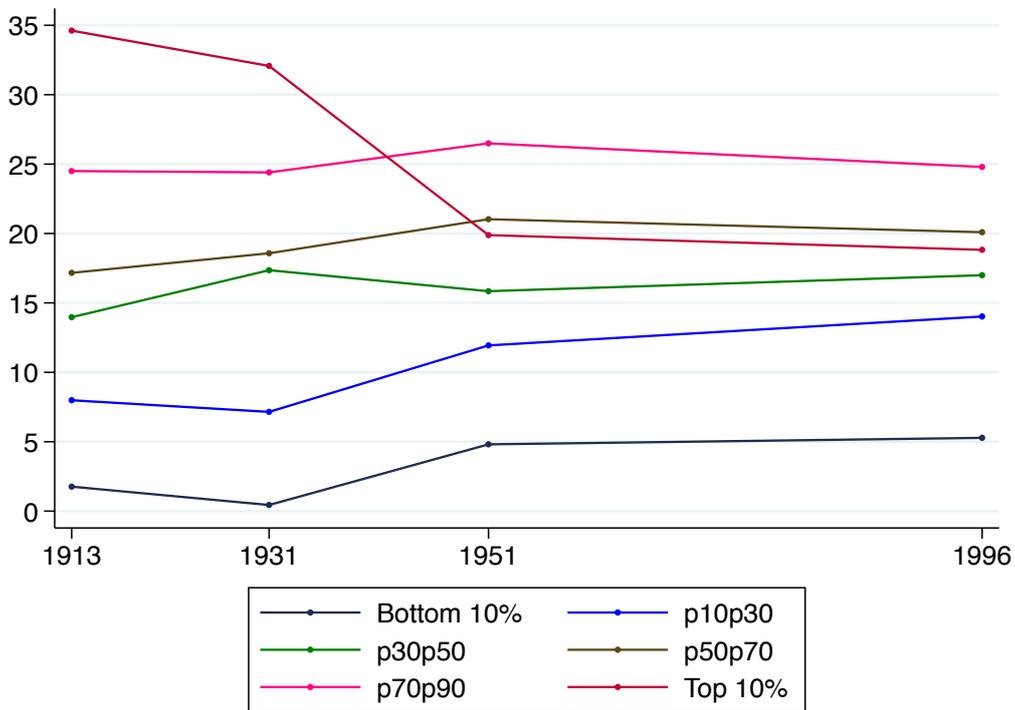


Notes: Graph 7 displays the ratios of the total of wages paid in each subgroup of the distribution (among civil servants) on the total wage bill (of civil servants). See Graph 1 for the definitions of subgroups.

Sources: Civil servants surveys and censuses, see Appendix D.

Graph 8 - Wage Shares for Civil Servants - Detailed Groups (1913 - 1996)

(in percentage of total wage bill)



Notes: Graph 8 displays the ratios of the total of wages paid in each subgroup of the distribution on the total wage bill, for civil servants. See Graph 2 for the definitions of subgroups.

Sources: Civil servants surveys and censuses, see Appendix D.

As can be seen on Graphs 7 and 8, civil servants experienced a clear decrease in wage inequality over the 20th century. The turning point seems to have occurred around World War II (the lack of data doesn't enable to know the precise timing of the evolution).

Until the years 1930, wage dispersion was clearly much more important in the public sector: the share of the top 10% was above 30% in 1913 and 1931, and the share of the top 1% was above 10%. Between 1931 and 1951, the shares of the top 10% and of the top 1% both decreased strongly (they respectively lost 12 and 9 percentage points). The top 10% share even became smaller than the share of the group between the 50th and the 70th percentiles (called p50p70 for simplicity). In 1951, both the top 10% and top 1% shares were below their counterpart from the private sector.

On the other hand, the share of the bottom 10% increased by 4 percentage points, as did the share of the group between the 10th and the 30th percentiles (p10p30), while the shares of p50p70 and p70p90 both increased by 2 percentage points. Overall, it was the bottom of the distribution whose average wage (compared to the general average) increased most between

1931 and 1951. Consequently, in 1951, the share of the bottom 10% was higher in the public sector than in the private sector.

Moreover, the reduction of the wage dispersion looks permanent, as the shares of 1996 were very close to the ones of 1951 (but the data is not detailed enough to display short-run fluctuations). Thus, since the middle of the 20th century, the public sector is no longer the driver of wage dispersion it used to be until World War II. This is all the more true as the weight of civil servants in the population of wage earners decreased over the second half of the century.

It is interesting to see that wage dispersion decreased in the public sector and not in the private sector. It reinforces the idea that wage inequality in the private sector hasn't decreased partly because it hasn't been fought against. Indeed, wage hierarchies have always been more visible in the public sector, because of their fixedness, transparency, and centralized character: this might have made them easier to denounce.

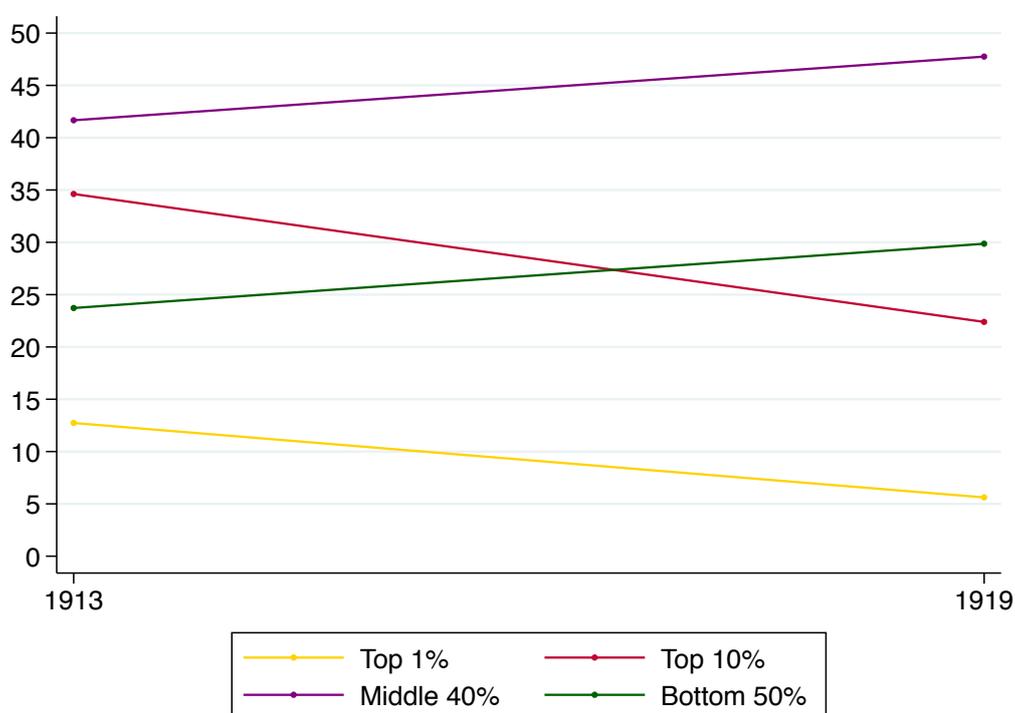
Short-run fluctuations

From 1919 to 1938: fluctuating wage inequalities

The impact of World War I

Graph 9 - Wage Shares - Main Groups (1913 - 1919)

(in percentage of total wage bill)

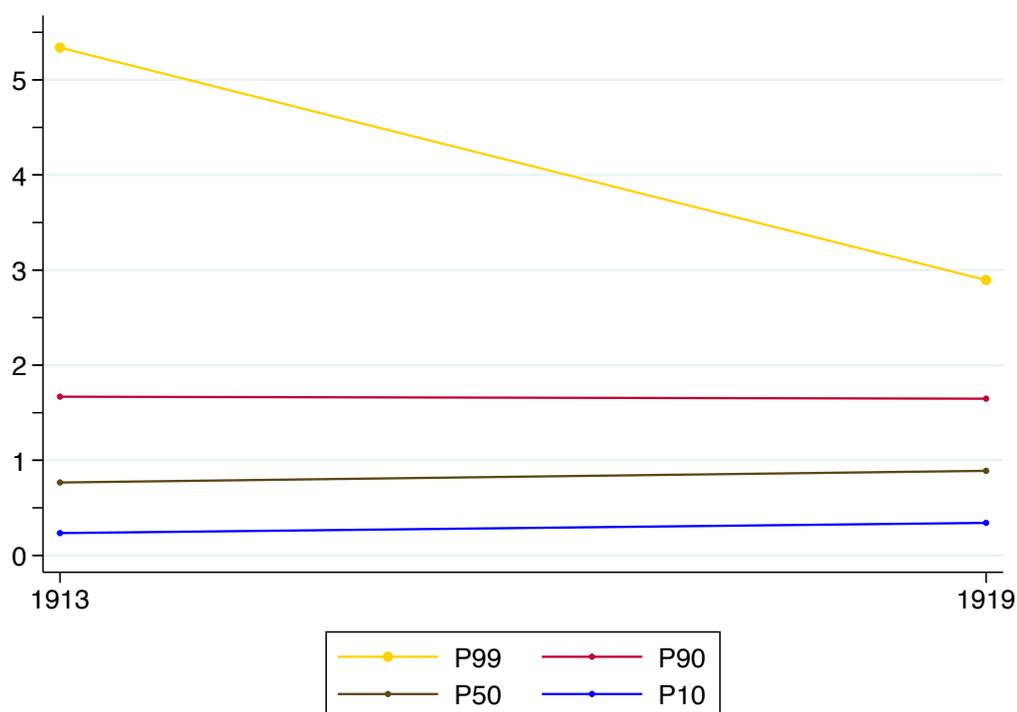


Notes: Graph 9 displays the ratios of the total of wages paid in each subgroup of the distribution on the total wage bill. See Graph 1 for the definitions of subgroups. For the year 1913, I used the distribution of wages for civil servants, considered to be representative of the distribution for all wage earners.

Sources: Civil servants survey for 1913 (see Appendix D), tax administration's summary statistics on the *Impôt cédulaire sur les salaires* for 1919.

Graph 10 - Percentiles - Main Groups (1913 - 1919)

(in proportion of average wage)



Notes: Graph 6 displays percentiles of the wage distribution, expressed in proportion to the national average wage. See Graph 5 for the definitions of percentiles. For the year 1913, I used the distribution of wages for civil servants, considered to be representative of the distribution for all wage earners.

Sources: Civil servants survey for 1913 (see Appendix D), tax administration's summary statistics on the *Impôt cédulaire sur les salaires* for 1919.

Assuming that civil servants can be considered as representative of the population of wage earners in 1913, Graph 9 enables to assess the impact of World War I on the wage distribution. Consistently to what was said in the existing literature, the First World war significantly reduced wage dispersion: the income share of the top 10% dropped from 35% to 22%. An important part of this drop (about 7 percentage points) can be imputed to the top 1%, which declined from 13% to 6%, suggesting that the very top of the distribution was even more affected by the war, (moreover P99 decreased more than P90, as can be seen on Graph 10). On the other hand the middle 40% of the distribution and the bottom 50% shares each increased by 6 percentage points.

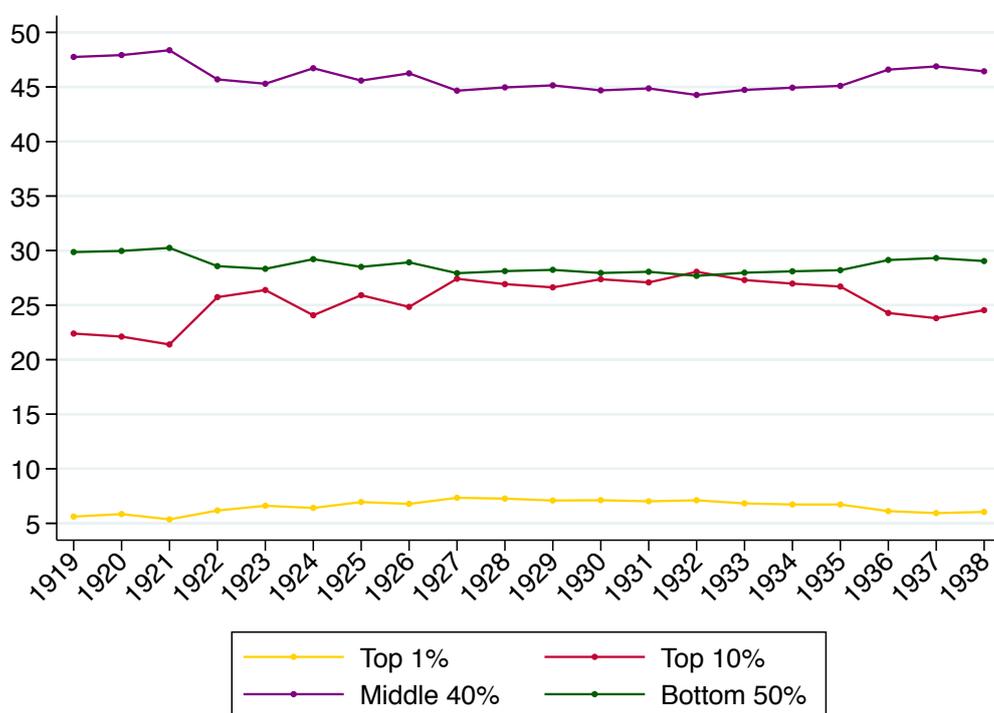
This is not surprising as wars are known to be times of compression of the wage distribution: top wages tend to be less protected than the rest from economic shocks triggered by wars, in particular inflation, so their average level (in constant terms) increases more

slowly than the rest. Consequently, the very low levels of the top 10% and top 1% shares in 1919 – 1920 can be seen as accidental.

From 1919 to 1933: the reconstruction of wage hierarchies

Graph 11 - Wage Shares - Main Groups (1919 - 1938)

(in percentage of total wage bill)

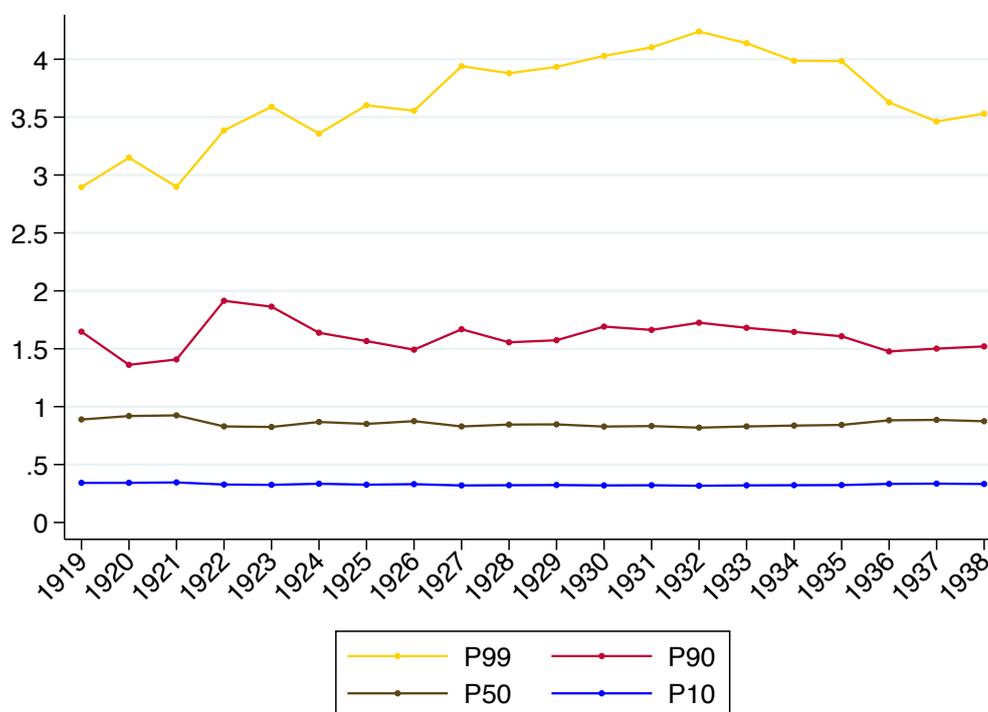


Notes: Graph 11 displays the ratios of the total of wages paid in each subgroup of the distribution on the total national wage bill. See Graph 1 for the definitions of subgroups. All sectors are covered from 1919 to 1938.

Sources: Tax administration summary statistics on the *Impôt cédulaire sur les salaires* (1919 – 1938)

Graph 12 - Percentiles - Main Groups (1919 - 1938)

(in proportion of average wage)



Notes: Graph 12 displays percentiles of the wage distribution, expressed in proportion to the national average wage. See Graph 5 for the definition of percentiles. All sectors are covered from 1919 to 1938.

Sources: Tax administration summary statistics on the *Impôt cédulaire sur les salaires* (1919 – 1938)

Despite regular oscillations (some of which might be overestimated because of the data⁹), there is a clear upward trend for the income share of the top 10% between 1919 and 1927 (from 22% to 27%), and a relative stabilization from 1927 to 1933.

This increase results from the post-war restoration of wage hierarchies. The political will to restore these hierarchies is exemplified by the successive revalorizations of civil servant wages: a first between 1919 and 1923, concerning the whole distribution, and a second between 1927 and 1931, focusing mainly on top wages, with a clear objective of returning to pre-war wage hierarchies. These are not only symbols, as civil servants were then an important part of the population of wage earners (one third of employees). Moreover, top wage earners didn't suffer from the crisis of the years 1930 as much as the rest. In particular,

⁹ The lowest percentile captured in the data source is the 91th in 1925, 94th in 1924, 95th in 1919 and 1922, 96th in 1923 (for other years, it is below the 90th).

they didn't experience partial unemployment, whereas an important number of factory workers did, hence the stabilization of the top 10% share around 27% until 1933.

This is all the more true for public workers, whose jobs and nominal wages were fixed, and who benefited from the deflation triggered by the crisis, whereas even private top wage earners suffered from the deflation at some point. The data doesn't display the fluctuations of the distribution of wages for civil servants between 1913 and 1931, but we can see on Graph 7 that in 1931, the top 10% share of civil servants had almost recovered its 1913 level (it was 35% in 1913 and 32% in 1931), despite a strong compression of the wage distribution in the public sector during the war.

The parallelism between the evolution of the middle 40% and the bottom 50% shares is artificial, as I imputed to each the share of the 90% they had in 1947. But it can be noted, in favor of this assumption, that in 1931 it yields shares of the bottom 90% very similar to the ones for civil servants, and the two distributions are known to be close during the interwar period. According to this assumption, the middle 40% decreased by 3 points and the bottom 50% by 2 points between 1919 and 1933, as a result of the restoration of wage hierarchies.

If we compare the distributions for the public sector only and for the entire population of wage earners in 1931, we can see that the share of the top 10% is significantly higher (by 5 points) in the public sector. This could be because the majority of very qualified wage earners still worked in the public sector during the interwar period, and so very high wages concentrated there. It is also possible that top wage earners in the public sector (university teachers, workers in ministries for instance) had a political influence that enabled them to protect their high remunerations. The shares of the bottom 50% and the middle 40% within the bottom 90% are very similar, as we said: respectively 62% and 38% for the whole population, 63% and 37% for civil servants.

It thus seems that the main difference between the public sector and the whole population concerns the top: the public sector was the main driver of top wages during the interwar period. However, the data probably underestimates inequality in the private sector, especially during the crisis, because it only takes into account full-time workers. Consequently it doesn't capture the impact of partial unemployment on the wage distribution, an important consequence of the crisis in the private sector.

Looking more precisely on Graph 8 at what happened within the bottom 90% for civil servants, there seems to be a clear demarcation at the 30th percentile: p10p30 and the bottom

10% both decreased compared to their 1913 levels, whereas p30p50 and p50p70 increased, and p70p90 stayed constant. So the reconstruction of wage hierarchies in the public sector seems to have benefited the middle of the distribution as well (apparently even more than the top 10%).

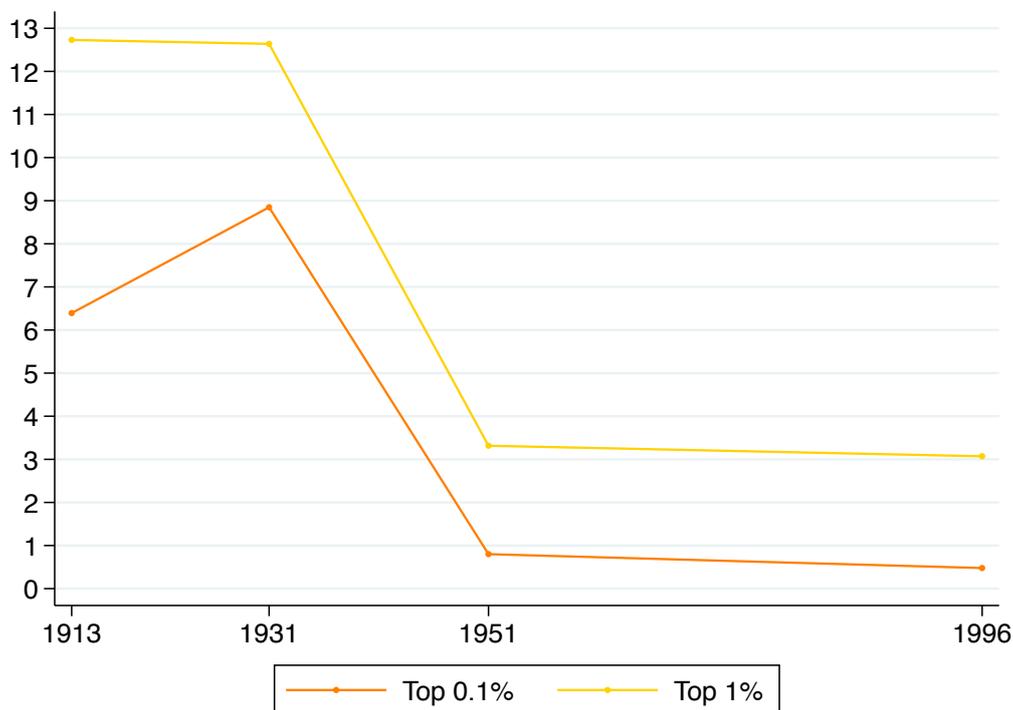
Graphs 11 and 12 enable to look closer at what happened within the top 10% for the whole population. The rise in the top 10% share seems in large part driven by the top of the decile: P99 increases more than P90 for instance. But there seems to be a slight difference between the public sector alone and the whole population in this respect.

For civil servants, the rise of the top decile looks mostly driven by the rise of the top centile: as can be seen on Graph 7, the top 1% share had completely recovered its 1913 level in 1931, whereas it was not the case for the top 10%, and the share of the top 0.1% increased (as can be seen on Graph 13 below).

For the whole population, the rise of the top 1% share only accounted for about 1.5 percentage points of the overall rise of the top decile between 1919 and 1927. This is consistent with the fact that the share owned by the top 1% was much higher for civil servants. Thus the bottom 9% of the top decile benefited more from the restoration of wage hierarchies in the private sector, whereas the benefits were apparently captured by the top 1% in the public sector.

Graph 13 - Wage Shares within the Top 1% for Civil Servants (1919 - 1996)

(in percentage of total wage bill)



Notes: Graph 13 displays the ratios of the total of wages paid in each subgroup of the distribution on the total wage bill, for civil servants. The Top 1% gathers individuals whose wage is above (or equal to) the 99th percentile, the Top 0.1% gathers individuals whose wage is above (or equal to) the 99.9th percentile.

Sources: Civil servants surveys and censuses, see Appendix D.

From 1933 to 1938: new wage hierarchies challenged

As can be seen on Graph 11, the top 10% share declined from 27% in 1933 to 24% in 1937, with a clear drop in 1936, so that by 1938 it was at a similar level than in 1922. Consequently, the shares of the bottom 50% and of the middle 40% respectively increased by 1 and 2 points. Again, this decrease is partly, but not totally, driven by the top: the share of the top 1% decreased by 1 point.

This decline first occurred because the deflation triggered by the crisis also affected top wages in the private sector, even if more moderately and later than middle and bottom wages. Moreover, in 1936 the newly elected *Front Populaire* revalorized only factory workers wages. So top wage earners, especially in the public sector, were more affected by the inflation that followed the election of the *Front Populaire*. (The slight increase of the top 10%

share in 1938 could be due to the revalorization of civil servant wages that occurred that year).

Consequently, at the eve of World War II, wage inequalities had only slightly increased compared to what they were just after World War I.

From the end of World War II to 1968

The impact of World War II

The data doesn't display the drop in top shares that occurred during World War II, according to the literature (Piketty 2001, 2014). Like during World War I, low wages were more generously protected against inflation, so top shares decreased.

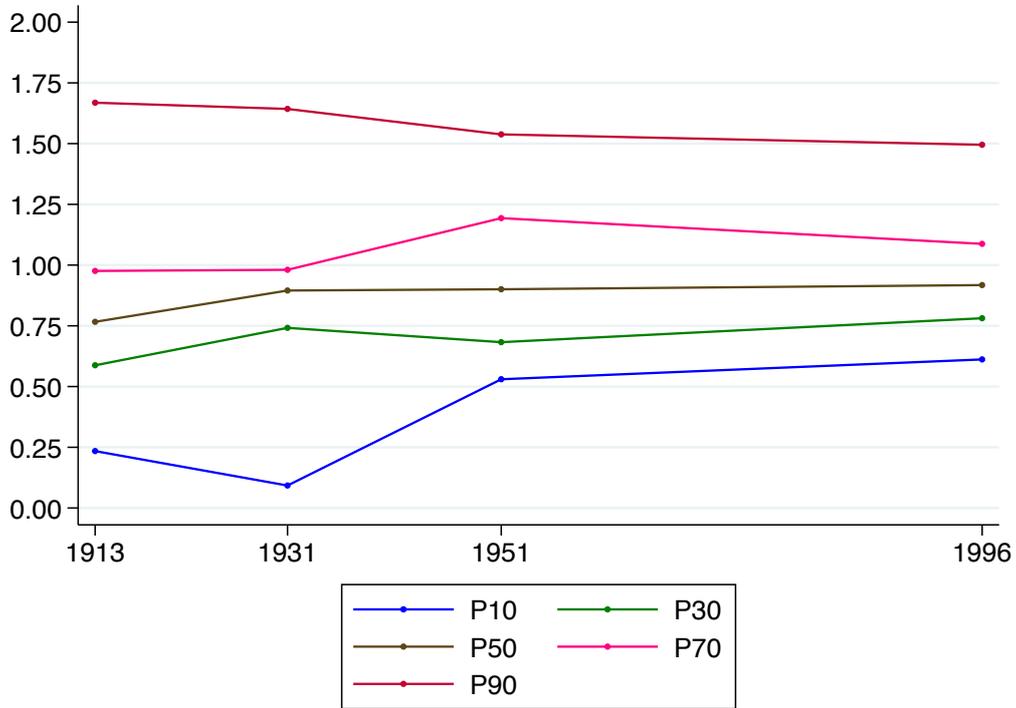
Instead, Graph 3 displays an increase in the top 10% share (of 2.5 points) from 1938 to 1947. This is because as soon as 1945 there was a strong political will to quickly restore pre-war wage hierarchies, embodied in the "arrêtés Parodi". It appears to have been efficient, as the top 10% share recovered its pre-1936 level as soon as 1947, and the top 1% share was slightly higher in 1947 than in 1938.

On the other hand, as previously developed, civil servants experienced a very different evolution. The top 10% and top 1% shares decreased strongly. P90 (expressed in proportion of average wage) decreased only slightly (from 1.6 to 1.5), as can be seen on Graph 14, which suggests that the decline mostly occurred at the top. But the bottom 9% of the top decile was also affected (its income share lost 3 points). The very top was also much affected: as can be seen on Graph 13, the top 0.1% share was divided by 10 between 1931 and 1951. This evolution mostly benefited the bottom of the distribution: both the bottom 10% and p10p30 saw their share increase by 4 points, and P10 increased by 0.4.

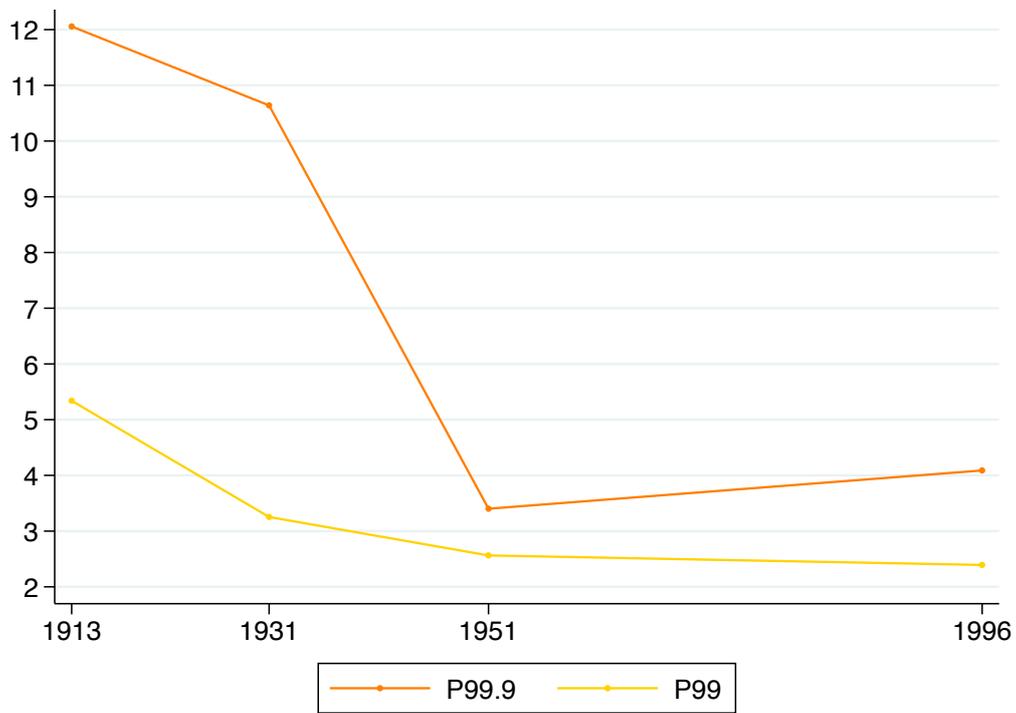
Graph 14 - Percentiles for Civil Servants (1913 - 1996)

(in proportion of average wage)

Graph 14a – Detailed Groups



Graph 14b – P99 and P99.9



Note: Graph 14 displays percentiles of the wage distribution for civil servants, expressed in proportion of the average wage (of civil servants). P99 is the 99th percentile, P99.9 is the 99.9th percentile.
Sources: Civil servants surveys and censuses, see Appendix D.

Consequently, there was a complete reversal of hierarchies between the public and the private sectors: the top 10% and top 1% shares were in 1947 much higher in the private sector (26% and 6.5% respectively), whereas the share of the bottom 10% was much lower there (5% for civil servants and 3% in the private sector).

The data doesn't display the precise timing of this reversal. The fact that public wage upgrades during World War II mostly focused on low wages probably played a role. After the war, measures were adopted to restore high wages in the public sector, but they were all much more modest than before: the new civil servants pay scale of 1948, significantly more compressed than previous ones, embodies that. Overall, World War II seems to have marked an end to very high wages in the public sector, and to civil servants driving wage inequalities.

The end of the years 1940 is a period of strong inflation. As higher wages are usually less protected from inflation, this can explain the decrease of the top 10% share between 1947 and 1950 (that can be seen on Graph 4).

From 1950 to 1968: a steady increase in wage inequalities

There are several versions of the detailed fluctuations of the years 1950, depending on the precise corrections applied from 1953 to 1962 (details are in Appendix B). Whatever the version used, the top 10% share increased significantly from 27% in 1952 to 31% in 1963 (both non corrected years), and then stagnated at its high level until 1967 – 1968, as can be seen on Graph 4. Overall, it increased by 4.5 percentage points between 1950 and 1967.

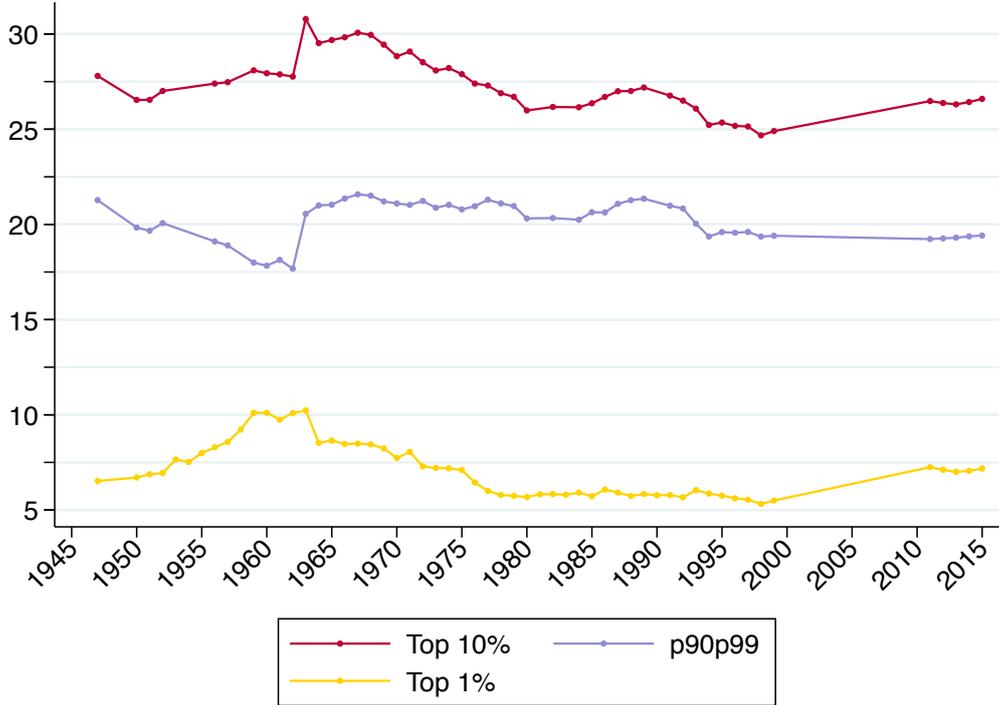
On the other hand, p50p70 and p30p50 lost 1.5 points, while p70p90 and p10p30 lost one point. The bottom 10% actually increased by one point, but mostly from 1950 to 1951, probably because of the implementation of the minimum wage (the SMIG), and it stagnated from 1951 to 1968. So the widening of wage dispersion only benefited to the top, not even the upper-middle, while the bottom of the distribution (under the 30th percentile) seemed slightly better protected against it than the middle.

It is interesting to see that most of the increase of the top 10% share seems driven by the very top, at least until 1963. Indeed, 3 points of the 4-point increase in the top 10% share between 1952 and 1963 were imputable to the top 1%, as can be seen on Graph 15. This is

confirmed by the stability of P90 on Graph 5, suggesting that the bottom of the top decile wasn't much affected, whereas P99 slightly increased. However, from 1964 to 1968, the top 1% share started to decline, whereas the bottom 9% increased by one percentage point. This is confirmed by the Pareto coefficients of the top decile (b90) and the top centile (b99) on Graph 16, which increased until the beginning of the years 1960 and then decreased.

Graph 15 - Wage Shares within the Top 10% (1947 - 2015)

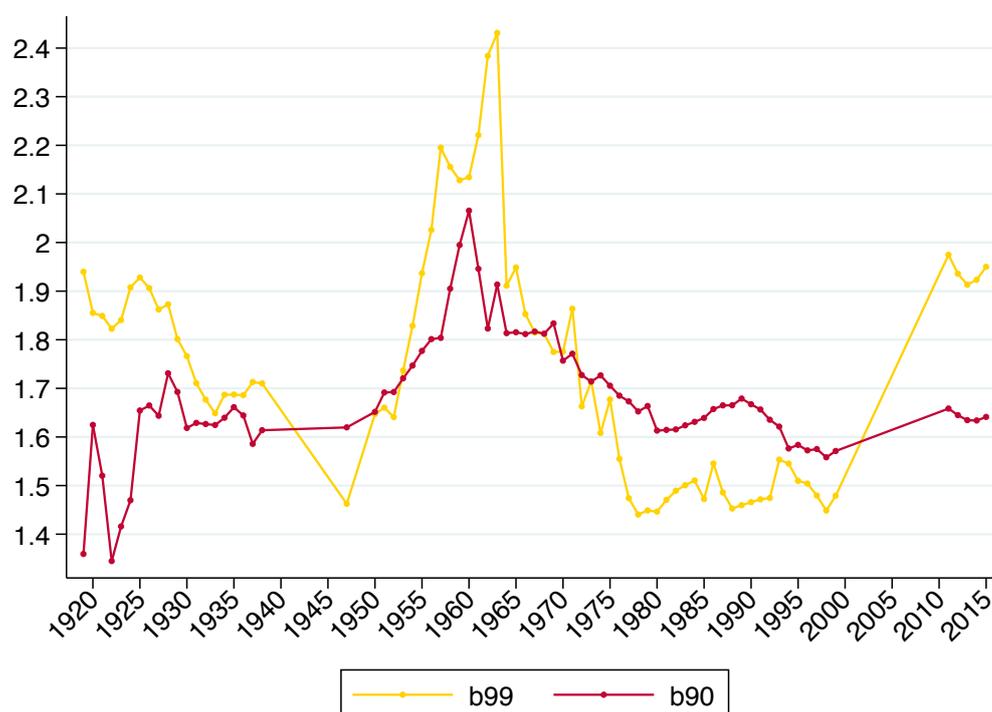
(in percentage of total wage bill)



Notes: Graph 15 displays the ratios of the total of wages paid in each subgroup of the distribution on the total wage bill, in the private sector. p90p99 gathers individuals whose wage is between the 90th and the 99th percentiles.

Sources: DADS

Graph 16 - Pareto Coefficients (1919 - 2015)



Notes: Graph 16 displays Pareto coefficients. b99 is the ratio between the average wage of the top 1% of the distribution and the 99th percentile, and b90 is the ratio between the average wage of the top 10% of the distribution and the 90th percentile. All sectors are covered from 1919 to 1938, only the private sector is between 1947 and 2015.

Sources: Tax administration summary statistics on the *Impôt cédulaire sur les salaires* (1919 – 1938), *Déclarations Annuelles de Données Sociales* (DADS) (1947 – 2015).

Meaning: In 1963, the average wage of the top 10% of the distribution is equal to 1.91 times the 90th percentile.

Overall, the significant increase of the top 10% share was at the beginning mostly driven by the very top, but the bottom of the top decile took over at the beginning of the years 1960. However, the findings regarding the top 1% must be interpreted with caution: they result from a threshold interpolation with a given national average in the input data, and in that case the generalized Pareto interpolation algorithm “fills” the gap between the average it computes and the given average by modifying the top of the distribution. So it is not impossible that such a differential evolution between the top 10% and the top 1% share is somewhat artificial.

These findings are entirely consistent with the literature, and they confirm that even the upper-middle part of the distribution didn’t follow the top 10%. They also show that the repartition within the bottom 90%, even between small subgroups of the distribution, remained very stable over the period, despite the transformations of the labor market occurring then.

This exclusive rise of the top of the distribution can partly be explained by the theory of supply and demand of qualifications. In a context of reconstruction and then modernization of the economy, the demand for qualified work was high, and the general increase of the labor force's qualifications hadn't borne fruit yet. So, in the years 1950, very qualified workers such as engineers and other "cadres supérieurs" still concentrated at the very top of the distribution, and their qualifications became very much in demand, and thus even better remunerated. Indeed, the ratio between the average wage of a "cadre" and the average wage of a factory worker ("ouvrier") increased over the period, whereas the ratio between a factory worker and an employee remained constant.

The catching up of the bottom 9% of the top decile could then be explained by the increase of the number of "cadres supérieurs": not big enough to make their remunerations fall, but big enough to affect people at the bottom of the top decile. Besides, the bottom 9% might also have more suffered from the strong inflations of 1951-1952 and of 1956-1957 than the top 1%, explaining their belated catching-up.

The reduction of wage inequality was also clearly not the priority of the successive governments of the period. Their focus in terms of economic policy moved from reconstruction to modernization to international competitiveness, (with the signature of the Treaty of Rome in 1957 creating the European Economic Community). This is embodied by the evolution of the minimum wage, created in 1950: it was almost never revalorized to follow inflation. Furthermore, the economic activities on which growth relied (metal industries for instance) needed cheap labor, and this could have provided an additional disincentive to protect low wages. The figure of the immigrant workers of the years 1950 and 1960, known to be the essential cheap labor force behind the growth of the "Trentes Glorieuses", is a significant symbol in this respect for instance.

The strong increase of wage inequalities in the private sector might be one of the reasons of the very poor reputation of economic growth during the years 1960, and for the growing social discontent that exploded in May 1968.

From 1967 to 1980: the catching up of low wages

As can be seen on Graph 4, the top 10% share declined steadily from 30% in 1968 to 26% in 1980, and then stagnated until 1983. This decline concerned only the top 10%: both p70p90

and p50p79 shares remained constant. On the other hand, the bottom 10% and p30p50 shares both increased by one point, while p30p50 share increased by 1.5 percentage point, and P10 (expressed in proportion of the average wage) moved from 0.4 to 0.5.

Within the top 10%, most of the decline can be imputed to the top: the share of the top 1% lost 3 percentage points between 1967 and 1980, while the bottom 9% decreased by 1.5 points (Graph 15). Besides, P90 remained relatively constant whereas P99 decreased by 0.5 point (Graph 5), and the Pareto coefficient of the last decile, b90, moved from 1.8 to 1.6 (Graph 16). All these indicators suggest a relative compression of the last decile. It can also be noted that the top 1% share started and stopped to decrease earlier: the bulk of its decline occurred between 1963 and 1978, while the bottom 9% decreased substantially only between 1977 and 1980.

The main reason for the decline in the top 10% share is the strong revalorization of the minimum wage. Following the social upheaval of May 1968 in France, the “Accords de Grenelles” of May 1968 increased the minimum wage by 20%. It also became partially indexed on factory workers hourly average wage (the TSH¹⁰), not only on prices, and was much more often revalorized by successive governments over the whole period. Overall, its purchasing power increased by 130% between 1968 and 1983, against only 50% for the average wage. It even progressed quicker than industry workers wages: it increased by 5.8% per year whereas the TSH increased by 3.8% per year.

The rise in the minimum wage apparently spread to the whole first half of the distribution, as can be seen on Graph 4. It didn't affect the top, which continued to progress but at a more moderate rhythm, resulting in the decline of the top 10% share mentioned above. Besides, the years 1976–1977 are known to be a turning point for wages: their progression markedly slowed down from that point onwards, whereas the minimum wage was not affected until the years 1980.

On the other hand, the rise in the qualification level of the labor force, that really took place in the years 1960, could explain both why the share of bottom 9% of the top decile didn't decline much before 1977, and why the share of the upper middle of the distribution stayed constant. Maybe the wage bonus associated with being a qualified worker (“cadre supérieur”) was spreading below the 99th centile and even below the top decile, making the

¹⁰ The “taux de salaire horaire ouvrier” is the average of the hourly wage of different categories of factory workers (so it is not equal to the average wage of factory workers, which depends on the number of hours worked and of the structure of the working population).

average wage of these groups grow faster than in the top 1%, and thus keeping their share constant. It could be that this phenomena stopped at the end of the years 1970 for the bottom 9% of the top decile, as it had for the top 1% earlier, triggering a decline in their income share. On the other hand the top 1% was maybe less affected by the general slow-down in the progression of wages of 1977 - 1978, and so their share stopped to decrease.

From 1983 to the end of the 20th century

The period of time between 1983 and 1999 is mostly characterized by the stability of all income shares. Two phenomena can be noted nonetheless.

The top 10% share increased by one percentage point between 1984 and 1989 and fell again by two points between 1989 and 1994, before stagnating at a lower level than in 1980 (25%) for the rest of the years 1990 (Graph 4). This was apparently entirely driven by the bottom 9% of the top decile, which increased by one point between 1989 and 1994, whereas the top 1% share remained extremely stable throughout the whole period (Graph 15).

All other shares remained particularly stable. Only the bottom 10% share rose from 3% in 1991 to 4.6% in 1999.

Indeed, 1983 is known in France as the “*tournant de la rigueur*”, when the socialist government of François Mitterrand, after two years of moderately expansionist policies, turned towards austerity and disinflation policies, joining the majority of European countries. An important component of this turning point was the end of the regular revalorizations of the minimum wage, so the bottom of the distribution joined the general slow-down of wages. Furthermore, the moderation of wage progression was an integral part of the disinflation policies set up in response to the “*stagflation*” of the years 1970 (even nationalized companies participated to the general de-indexation of wages). It remained an important objective in the years 1990, as competitive disinflation policies were implemented in the context of the construction of the European monetary system, and more generally as an answer to the growing internationalization of production. Overall, this turn in economic policies joined the general slowdown of wages since 1976 to make wages stagnate throughout the entire distribution.

Thus, the small jump in the top 10% share at the end of the years 1980 could partly result from the oil counter-shock of the middle of 1985-1986 (assuming only high-wages benefited

from it), or from various measures of the Chirac government of 1986-1988. Fiscal measures, such as the suppression of the IGF (tax on high incomes) or a smaller rate for corporate taxes, might have created an incentive to increase high wages. It could also be that the strengthening of disinflation policies towards wages only affected the bottom 90%. Overall it is important to note that the explosion of wage inequalities that took place in the Anglo-Saxon world at the same period, linked to the rise of extremely high remunerations in certain sectors (corporate governance, finance), did not happen in France, at least before the end of the century.

As to the gradual rise of the bottom 10% share over the years 1990, it could be explained by the fact that upper parts of the distribution suffered from disinflation, while the minimum wage provided a limit to the slow-down of very low wages. However, it is important to note that the data includes only full-time workers, whereas taking into account the rise of part-time work that started in the years 1980 would most probably lead to an increase in wage inequalities, driven by the bottom of the distribution.

Evolutions in the private and the public sectors over the second half of the 20th century

Now, if we compare the evolutions within the private and the public sector over the second half of the 20th century (looking at Graphs 3, 4, 7 and 8), both distributions look overall very stable, and the differences that emerged after World War II between the private and the public sectors remained. A few divergences can be seen nonetheless.

Between 1951 and 1996, the top 10% share decreased by one percentage point in both sectors, however the top 1% share barely decreased for civil servants, whereas it lost one point in the private sector. This is maybe because the top 1% share had already dropped significantly in 1951 for civil servants. A similar remark can be made regarding the very bottom of the distribution: the bottom 10% remained almost constant for civil servants, whereas it gradually increased in the private sector (by 2.5 points between 1950 and 1996). This might be because the bottom 10% had already experienced a substantial increase between 1931 and 1951 for civil servants. As a result, the gap between the shares of the two sectors was maintained for the top 10% (25% in the private sectors and 19% for civil servants) but was reduced for the top 1% (5,5 % in the private sector and 3% in the public sector).

However, there was apparently more movement within the middle of the distribution for civil servants: while p70p90 and p50p70 decreased by one point, p30p50 and p10p30 respectively increased by one and two points. It looks as if after the general compression of the dispersion around World War II, only the interior of the distribution had moved, towards more equality as well. On the contrary, the shares of the middle of the distribution remained constant in the private sector, (except for p10p30).

Consequently, the shares of the bottom 10% were very similar in both sector, and the middle of the distribution became even more homogenous in the private sector compared to the private sector.

From 1999 to 2015: a new turning point for wage inequalities?

For the years between 1996 and 2014, using the *Enquête Revenus Fiscaux et Sociaux*, I was able to separate labor income earners into four groups: private sector wage earners, self-employed workers, public sector wage earners, and retirees and unemployed.

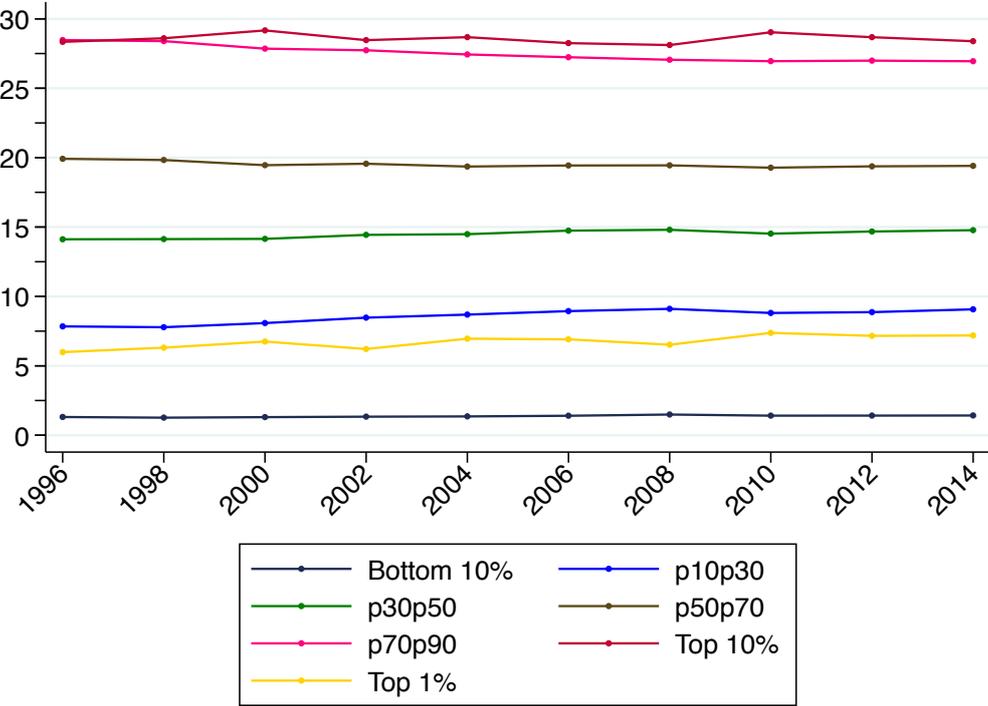
Labor income earners:

We first look at the distribution for all individuals of the sample whose labor income was superior to one third of the nominal minimum wage, the benchmark group.

The distribution looks remarkably stable: the top 10% share fluctuated between 28% and 29% over the whole period), while p50p70, p30p50 and the bottom 10% shares respectively stagnated around 20%, 14%, and 1.3 %. Only p70p90 and p10p30 seemed to follow a trend: p70p90 gradually decreased from 28 to 27%, and p10p30 increased from 8 to 9% until 2008 (then stagnated).

Graph 17 - Labor Income Shares for Labor Income Earners (1996 - 2014)

(in percentage of total labor income)

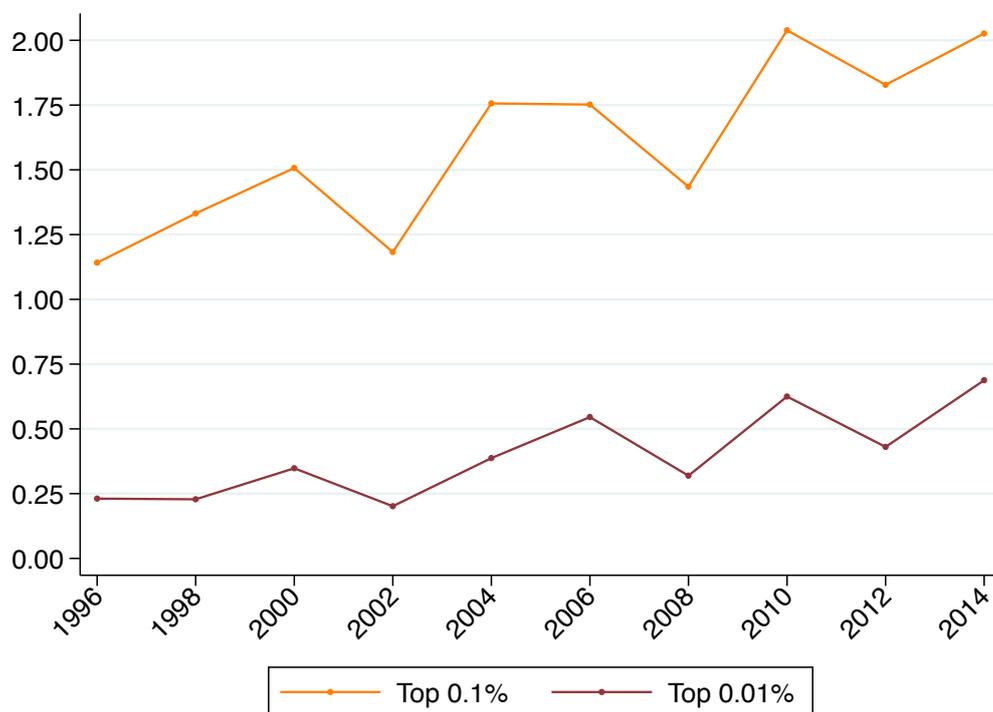


Notes: Graph 17 displays the ratios of the total of labor income earned in each subgroup of the distribution on total labor income, for positive labor income earners. See Graph 2 for the definitions of subgroups.

Sources: *Enquêtes Revenus Fiscaux et Sociaux (ERFS)*

Graph 18 - Top Labor Income Shares for Labor Income Earners (1996 - 2014)

(in percentage of total labor income)



Notes: Graph 18 displays the ratios of the total of labor income earned in each subgroup of the distribution on the total labor income, for labor income earners. The top 0.01% gathers individuals whose labor income is above the 99.99th percentile of the distribution.

Sources: *Enquêtes Revenus Fiscaux et Sociaux* (ERFS)

Graph 19 - Top Percentiles for Labor Income Earners (1996 - 2014)

(in proportion of average labor income)



Notes: Graph 19 displays the percentiles of the labor income distribution for positive labor income earners. See Graph 14b for the definitions of percentiles. P99.99 is the 99.99th percentile.

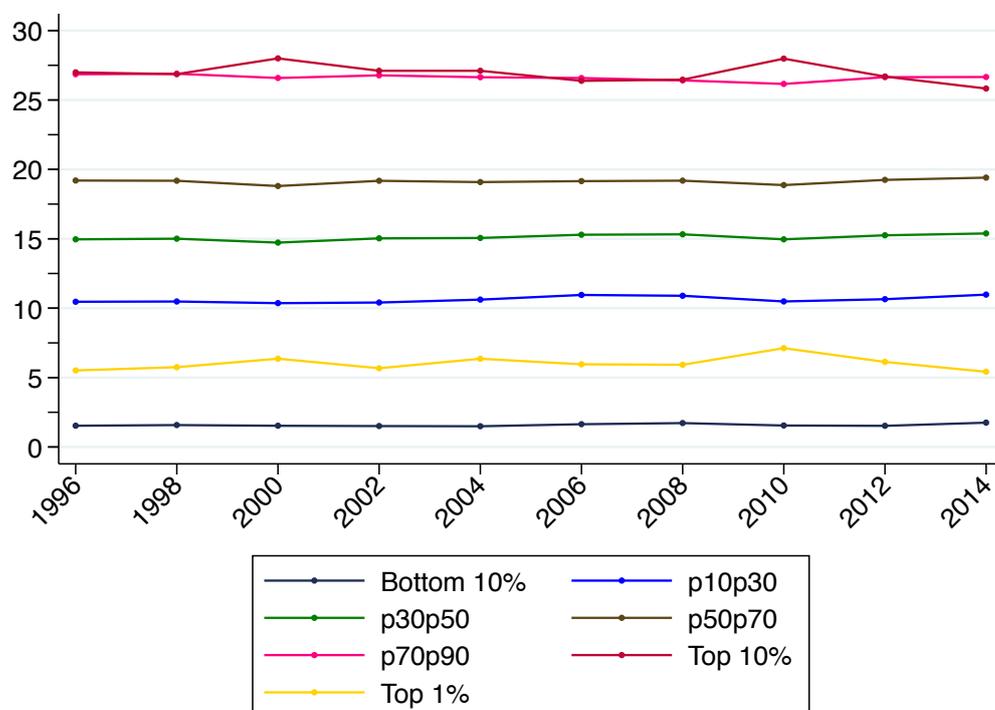
Source: ERFS

There seemed to be more movement at the very top: the top 1% share increased from 6% to 7%. This increase seemed to be driven by the very top: the top 0.1% share increased from 1 to 2%, and the top 0.01% share increased from 0.2 to 0.7%. This intuition is confirmed by the evolution of thresholds: P99 remained constant, while both P99.9 and P99.99 increased significantly (even more for P99.99).

Private sector wage earners:

Graph 20 - Labor Income Shares for Private Sector Wage Earners (1996 - 2014)

(in percentage of total labor income)



Notes: Graph 20 displays the ratios of the total labor income earned in each subgroup of the distribution (among private sector wage earners) on total labor income (of wage earners in the private sector). See Graph 2 for the definitions of subgroups.

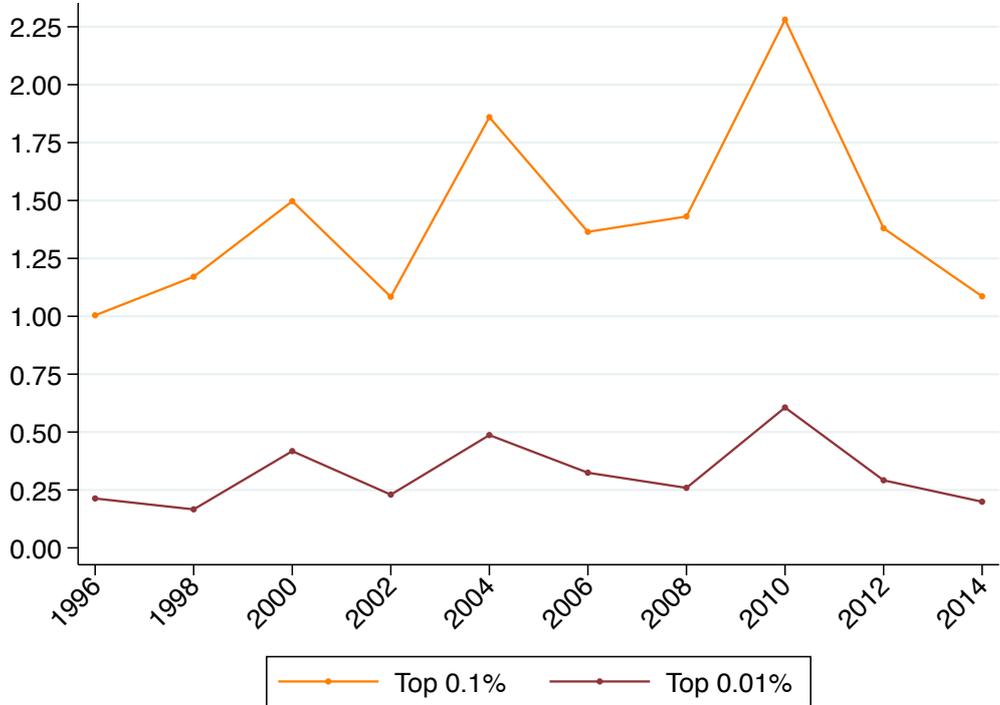
Sources: ERFS

As can be seen on Graph 20, the distribution of wages in the private sector was even more stable and narrower than for the benchmark group. The share of the top 10% fluctuated around 27%. A decreasing trend has maybe started in 2010, but this has to be confirmed. The shares of the various groups of the bottom of the distribution were very similar to ones of the benchmark distribution, but slightly lower for p50p70 and p70p90 (by one point) and slightly higher for the groups of the bottom 50% (by one point). There was no clear trend for any part of the distribution.

Within the top 10% of the distribution, the top 1% increased from 5.5% to 7% between 1996 and 2010 but decreased afterwards and came back to its initial level in 2014. The top 0.1 and 0.01 shares behaved in a similar way (see Graph 21): with important fluctuations, they increased significantly between 1996 and 2010 (respectively from 1% to 2% and from 0.2%

to 0.6%), but decreased after 2010 and came back to their initial level. However, according to DADS series (see Graph 3), the top 1% share remained around 7% between 2011 and 2015. The strength of DADS data compared to ERFS is its exhaustiveness. Thus, regarding the top of the distribution, DADS series might be more reliable than ERFS series, and so it can be supposed that the top 1% share did stagnate around 7% (but didn't increase further). Unfortunately, the significant decrease of the top 0.1% and 0.01% shares between 2010 and 2014 in ERFS series cannot be compared to DADS data.

Graph 21 - Top Labor Income Shares for Private Sector Wage Earners (1996 - 2014)
(in percentage of total labor income)



Notes: Graph 21 displays the ratios of the total labor income earned in each subgroup of the distribution on the total labor income, for wage earners in the private sector. See Graphs 13 and 18 for the definitions of subgroups.

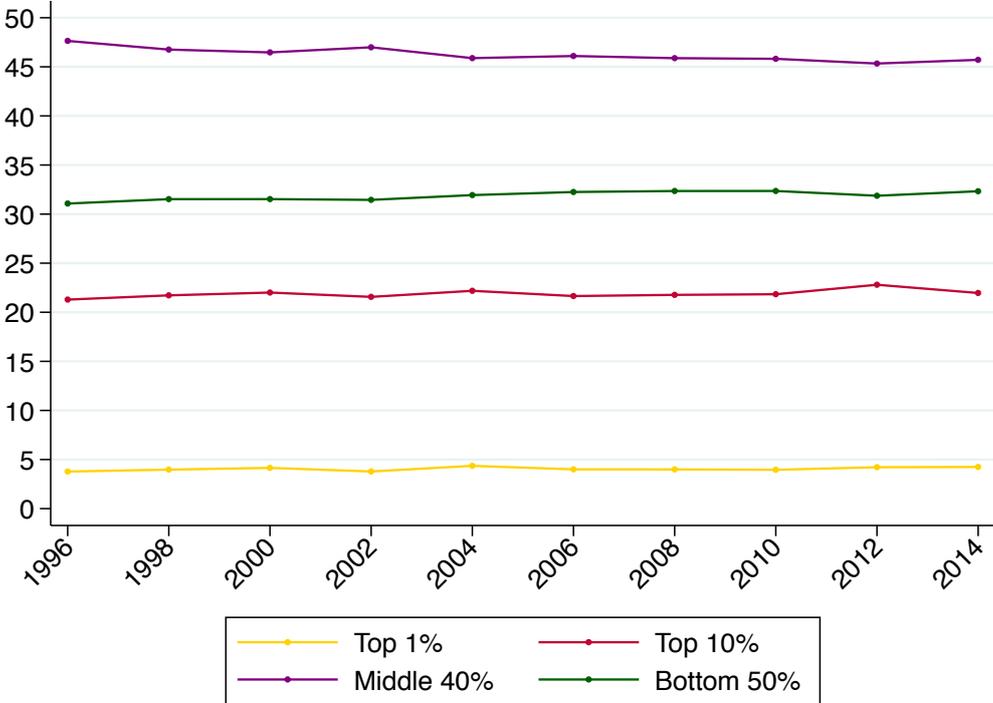
Sources: ERFS

Overall, there has indeed been an increase in very top wages over the years 2010. As for their potential decline since 2010, our sources disagree, but the pre-tax labor income series of Piketty, Garbinti and Goupille (2017) support the stagnation of the top 1% share displayed in the DADS series. If the drop of the top 0.1% and top 0.01% shares displayed in Graph 21 was reliable, it could be the result of the 2008 economic crisis. The question would remain

whether it is only temporary or whether it suggests that the rise of very high wages has come to and end.

Public sector wage earners:

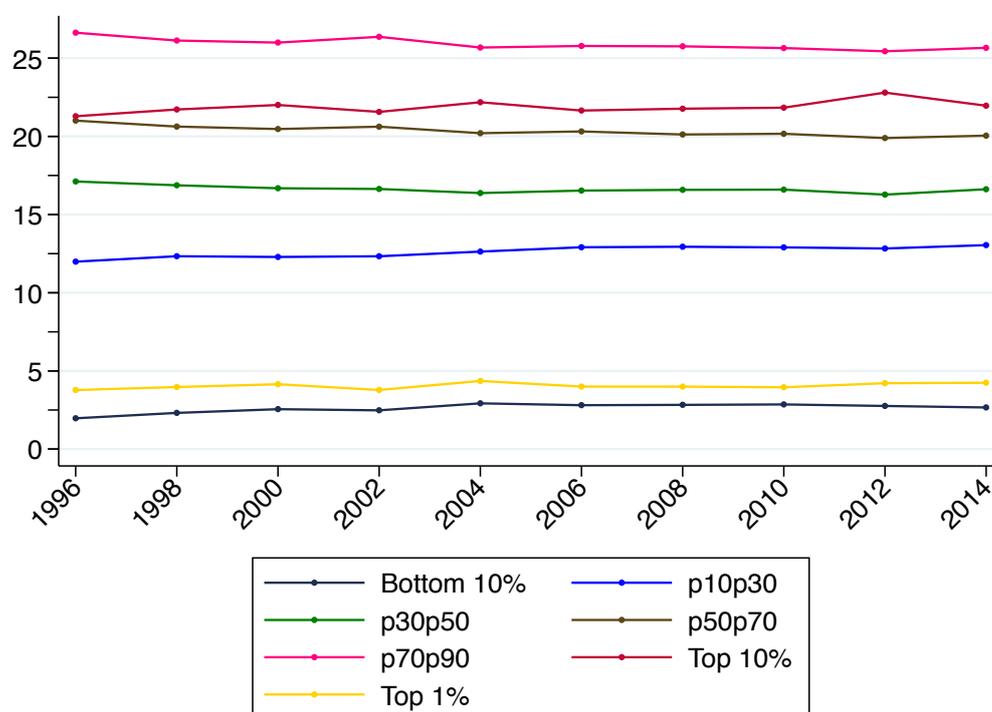
Graph 22 - Labor Income Shares for Public Sector Wage Earners - Main Groups
(1996 - 2014)
 (in percentage of total labor income)



Notes: Graph 22 displays the ratios of the total labor income earned in each subgroup of the distribution (among public sector wage earners) on total labor income (of public sector wage earners). See Graph 1 for the definitions of subgroups.
Sources: ERFS

**Graph 23 - Labor Income Shares for Public Sector Wage Earners - Detailed Groups
(1996 - 2014)**

(in percentage of total labor income)



Notes: Graph 23 displays the ratios of total labor income earned in each subgroup of the distribution on the total labor income, for wage earners in the public sector. See Graph 2 for the definitions of subgroups.

Sources: ERFS

Wages in the public sector were also remarkably stable (Graph 23), but displayed slightly more pronounced trends than in the private sector: overall, the share of the middle 40% fell from 48% to 46% between 1996 and 2014, while the share of the bottom 50% increased by one point, mostly thanks to p10p30. The share of the top 10% also increased by one point, but the top 1% remained stable.

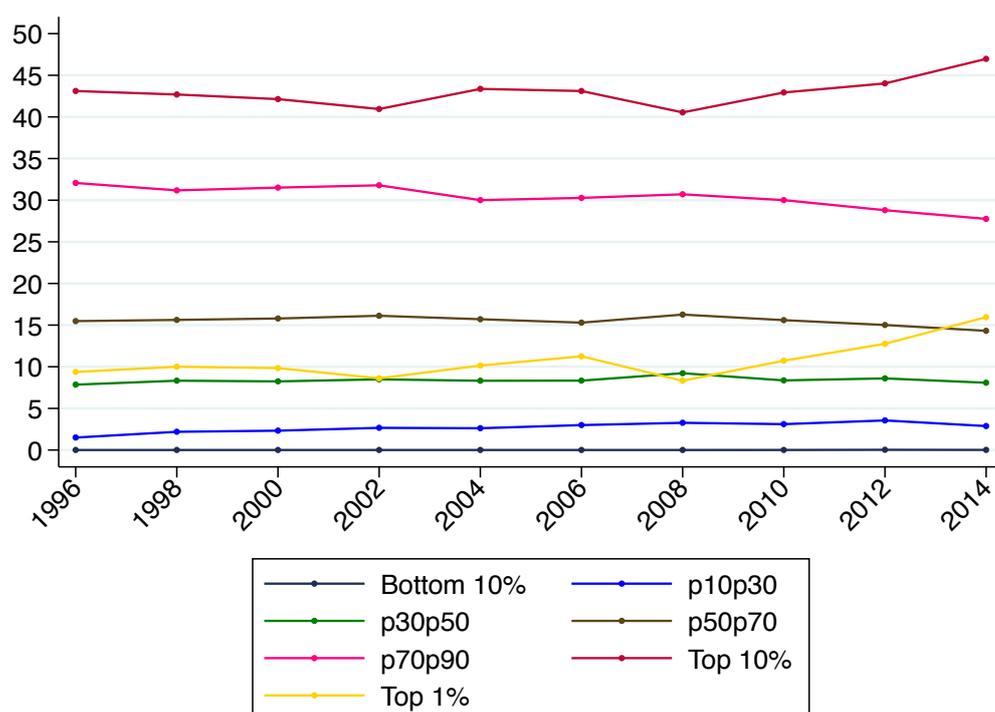
The wage hierarchy was much narrower than in the private sector, especially at the top: the top 10% share fluctuated around 22%, the top 1% around 4%, and p50p70 was roughly equal to 20% over the whole period. The middle of the distribution was quite similar to the private sector, but the bottom's shares were higher: around 12-13% for p10p30, and 2.5% for the bottom 10%.

Self-employed workers:

The distribution of labor income for self-employed workers is significantly different than for wage earners, both in terms of levels and evolution. As can be seen on Graph 24, the top 10% share experienced a significant increase over the period: after 12 years of fluctuations between 41% and 43% from 1996 to 2008, it steadily increased by 6 percentage point between 2008 and 2014, to reach 47% in 2014. On the other hand p70p90 decreased from 32% in 1996 to 28% in 2014. The rest of the distribution was relatively stable, though p10p30 increased by 1.5 points over the period. This widening of the wage hierarchy within the upper part of the distribution (the top 30%) was specific to independent workers.

Graph 24 - Labor Income Shares for Self-employed Workers (1996 - 2014)

(in percentage of total labor income)



Notes: Graph 24 displays the ratios of total of labor income earned in each subgroup of the distribution on total labor income, for self-employed workers. See Graph 2 for the definitions of subgroups.

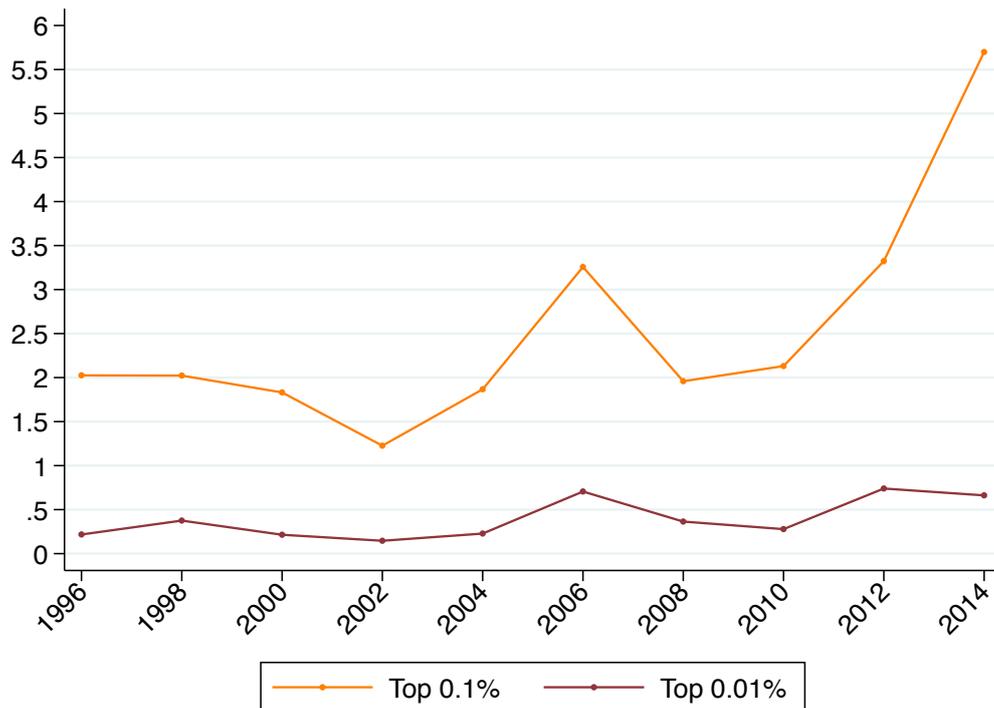
Sources: ERFs

Within the top 10%, the top 1% share displayed a similar (but even more impressive) evolution: it went from 8% in 2008 to 16% in 2014, which suggests it was the main driver of the increase of the top decile share. This is confirmed by the fact that P90 actually decreased

over the period. Within the top 1%, the top 0.1% share soared from 2 to 6% between 2008 and 2014 (Graph 25), whereas P99 increased only slightly (Graph 26): the increase of the top 1% share was itself driven by the top of the centile.

Graph 25 - Top Labor Income Shares for Self-employed Workers (1996 - 2014)

(in percentage of total labor income)

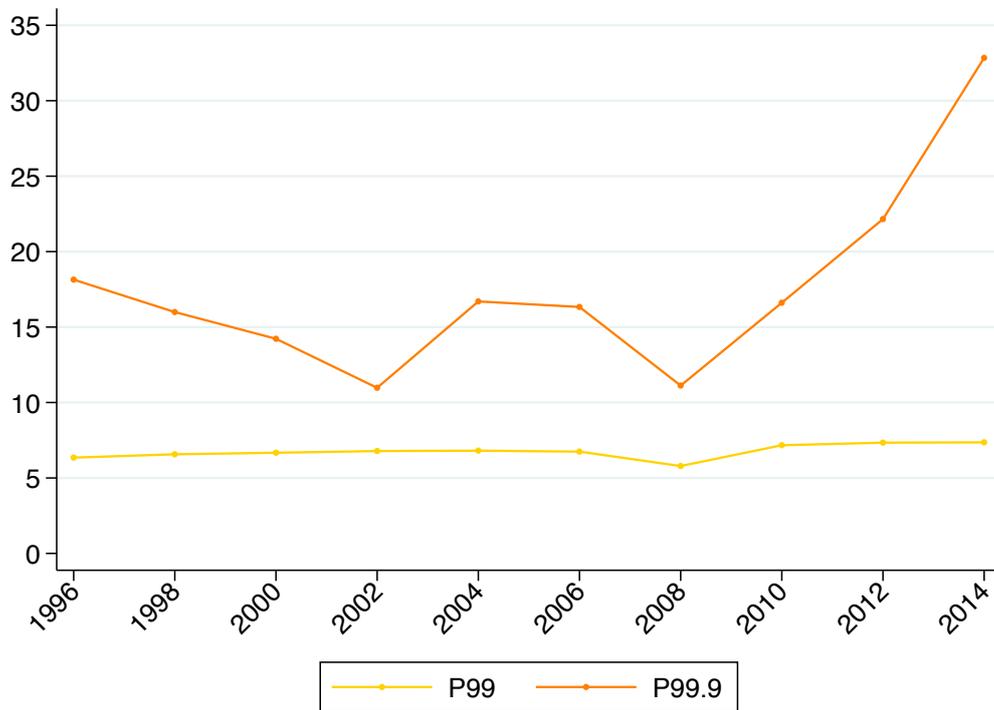


Notes: Graph 25 displays the ratios of total of labor income earned in each subgroup of the distribution on total labor income, for self-employed workers. See Graphs 13 and 18 for the definitions of subgroups.

Sources: ERFS

Graph 26 - Top Percentiles for Self-employed Workers (1996 - 2014)

(in proportion of average labor income)



Notes: Graph 26 displays the 99th and 99.9th percentiles of the distribution of labor income, for self-employed workers.

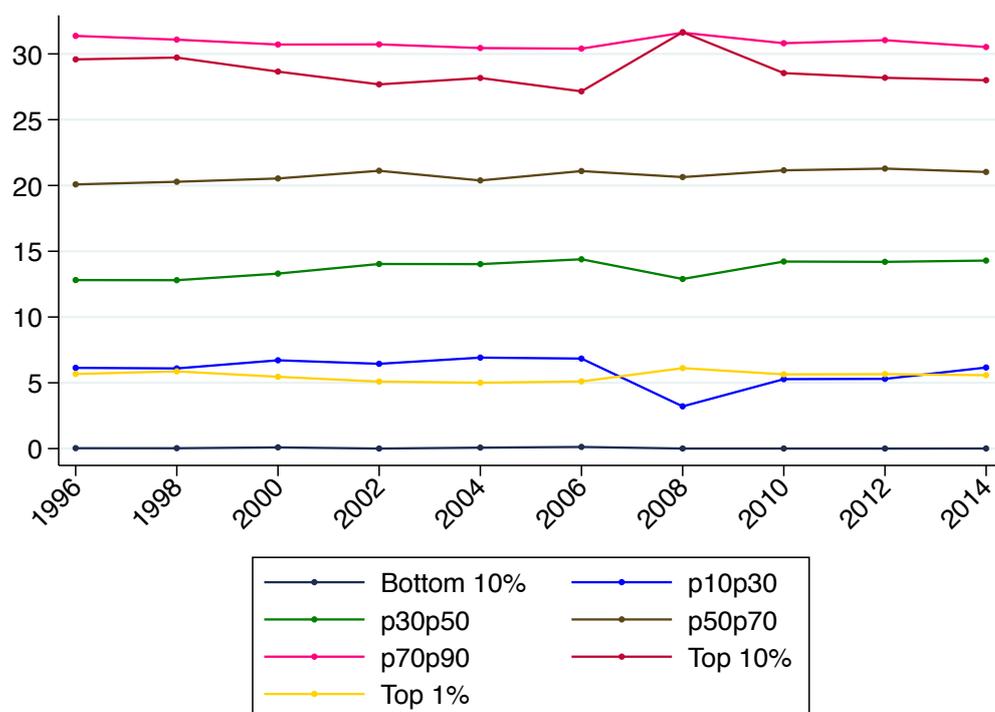
Sources: ERF5

In terms of levels, the distribution is much more spread out for independent workers than for the other groups, over the whole period. The top 10% share was always above 40%, so almost twice as much as in the public sector, and about 1.3 times its counterpart in the private sector (or in the benchmark group). It was much higher than the share of p70p90, but this last share itself was higher than for other sectors (about 30%). The top 1% share was also higher (around 10% for most of the period and it reached more than 15% at the end of the period). All the shares below the 70th percentiles were lower for independent workers, and the gap was the widest for p10p30 (2% for independent workers against around 10% for the other groups). It is possible that the data leads to overestimate these gaps: the number of people who don't currently work and have no income but call themselves "independent workers" in the *Enquête Emploi* might be higher than for than wage-earners, this is probably why the bottom 10% share is at 0.

Retirees and unemployed:

Graph 27 - Labor Income Shares for Retirees and Unemployed (1996 - 2014)

(in percentage of total labor income)



Notes: Graph 27 displays the ratios of total of labor income earned in each subgroup of the distribution (among retirees and unemployed individuals) on total labor income (of retirees and unemployed individuals). See Graph 2 for the definitions of subgroups.

Source: ERFS

The distribution of income for retirees and unemployed people is rather stable, as can be seen on Graph 27. The large drop in the share of p10p30 (and the corresponding increase in the share of the top 10%) in 2008 can probably be ignored, because these don't appear when we look at the distribution of pensions only (see Graph AE – 1 in Appendix E), so they are probably be linked to a specific feature of the questions related to unemployment and retirement in the *Enquête Emploi* in 2008.

The share of the top 10% of the distribution experienced a very slight drop of 1.5 points, matched by a corresponding increase in the shares of p30p50 and p70p90. However p10p30 decreased overall and the bottom 10% remains constant at 0. So overall, there was a very small narrowing of the distribution above the 30th percentile over the period.

In terms of levels, the distribution for retirees and unemployed is a bit more spread out than for the benchmark group: the top 10% was similar but p70p90 was above 30% throughout the whole period. The middle of the distribution was similar, but the bottom 30% had a lower share among retirees and unemployed. This might be slightly data-driven, as people with no current pensions could declare themselves retirees or unemployed, whereas observations in the benchmark group necessarily had an income above 0.

Overall the distribution of labor income among the population with positive labor income (the benchmark group) is close to the distribution of private wages, especially in terms of trends. This is not surprising given the proportion of private wage earners in the active population.

The distribution of labor income appears to have been stable in the years 2000, except for a significant rise of very top shares, both among self-employed workers and private sector wage earners.

Comparing ERFs and DADS

DADS and ERFs series for private wage earners

Focusing on private wage earners only, DADS and ERFs data are reassuringly similar. Noticeable differences concern the top and the bottom of the distribution: the top 10% share is slightly higher in ERFs data (27% instead of 25%) in 1996 and 1998, but similar (around 26%) in 2012 and 2014, and the shares of p70p90 and p50p70 are one point higher in ERFs series. On the contrary, the share of p10p30 is one point lower in ERFs, and the bottom 10% share is much lower (in ERFs), especially at the end of the period (1.7% against 3.9% in 2014). The more widely spread ERFs distribution can be explained by the inclusion of part-time workers in ERFs data. DADS didn't include part-time workers before 2009, and then they were included as "équivalents temps plein" (their wages were upgraded as if they were full-time workers). Higher top shares could also result from the fact that secondary sources of income are taken into account in the ERFs data.

However, the top 1% share was higher in the DADS data at the end of the period (around 7% versus 5.5 % in the ERFs data). This is probably because DADS data captures top wages better, because its sample is exhaustive, whereas the ERFs sample is based on the *Enquête Emploi*, and surveys are known to underestimate incomes at the top.

Despite these slight differences, the two series can be considered close.

DADS and ERFs series for labor income earners

The same remarks can be made about the differences between the DADS series and the ERFs series for the benchmark group, since this distribution is even more widespread than the distribution for private wage earners. But the differences are not huge, even if they are significant: 2 to 3 percentage points for the top 10% and p70p90, barely one point for p50p70 and p30p50. They are more problematic at the bottom: 4 to 3 points for p10p30 (the gap decreased over time), and almost 2.5 points for the bottom 10%.

Extending the ERFs series until 1970 would enable to see how these gaps vary over time, and to understand how DADS series could be corrected to match better the whole population of labor income earners.

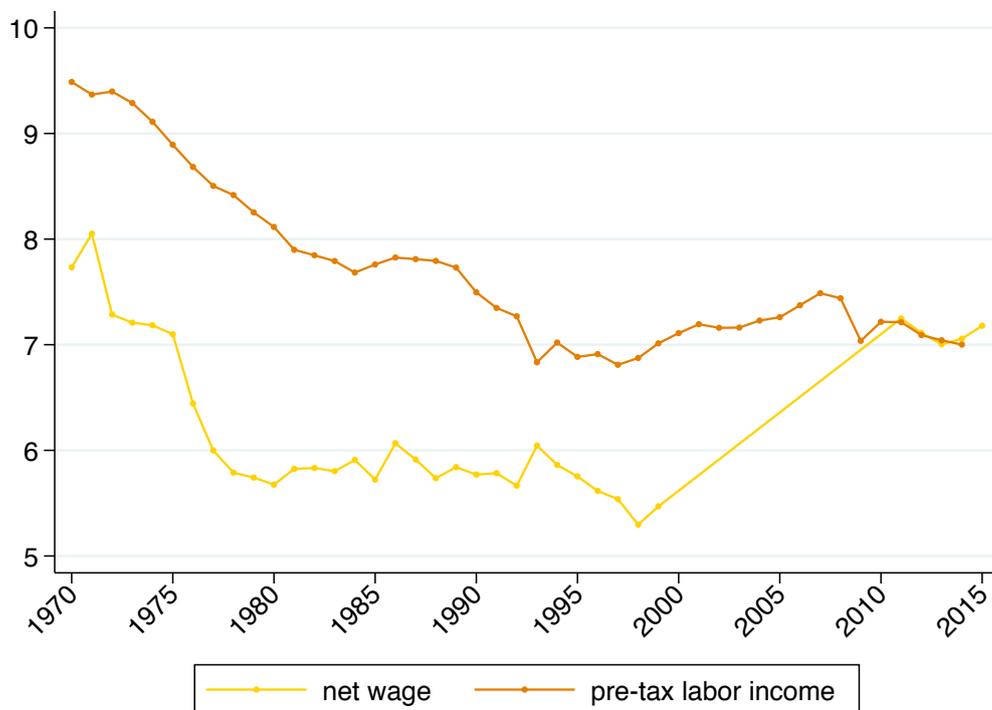
Comparing net wage and pre-tax labor income series

Comparing the DADS series with the series in pre-tax labor income constructed by Piketty, Garbinti and Goupille (2017) on Graph 28, one can see that there are important gaps between net wage and pre-tax income shares for all subgroups.

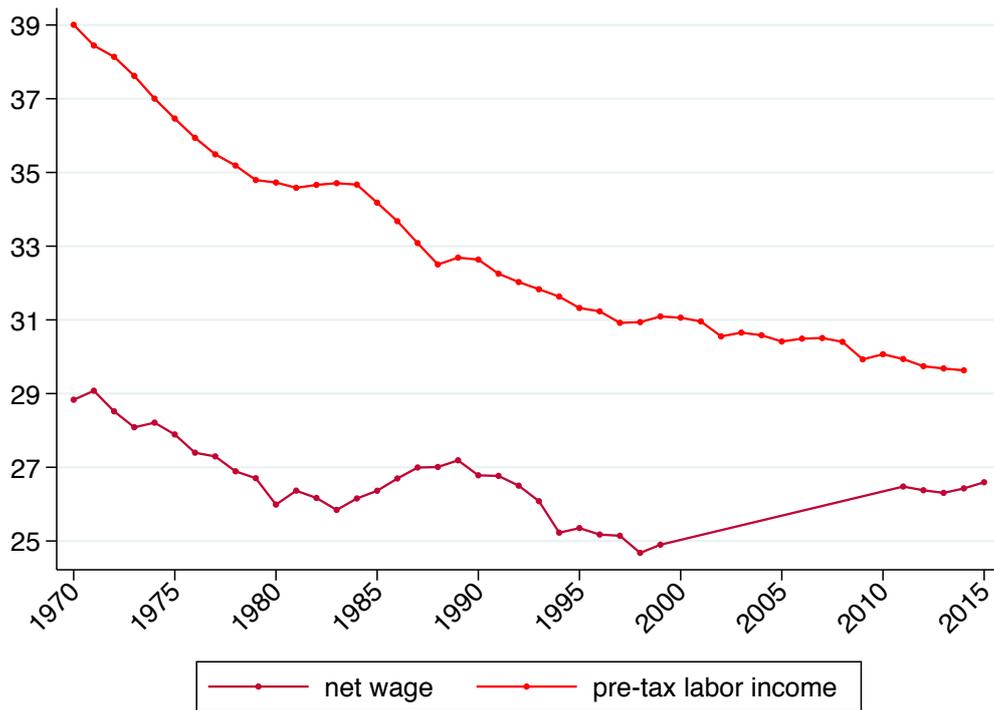
Graph 28 - Labor Income Shares in Net Wages and Pre-tax Labor Income (1970 - 2015)

(in percentage of total net wage or labor income)

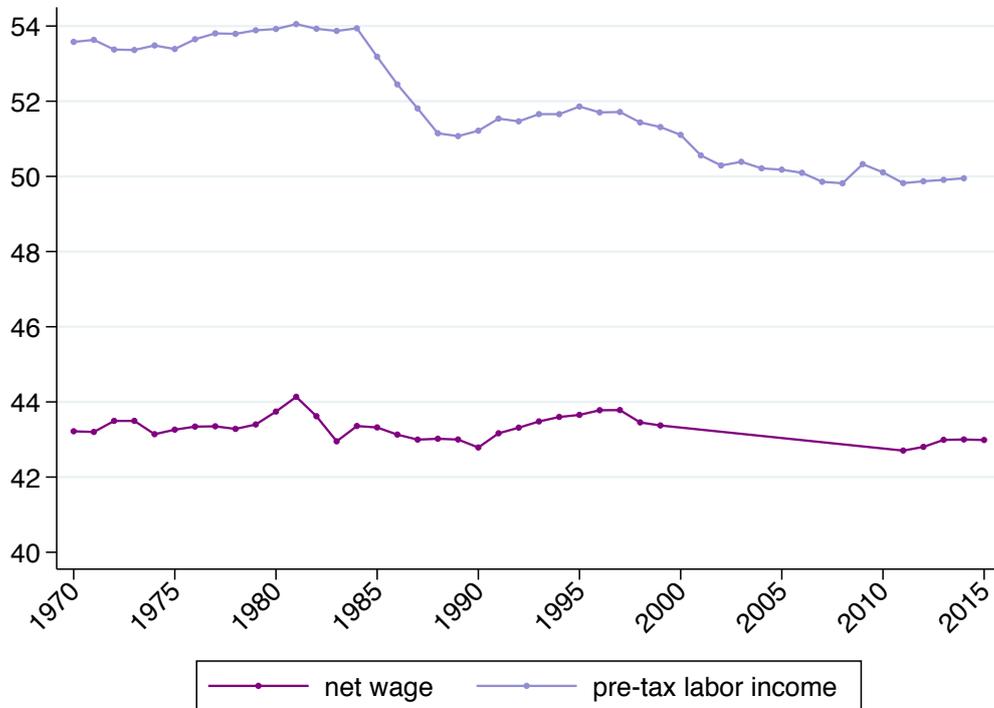
Graph 28a – Top 1%



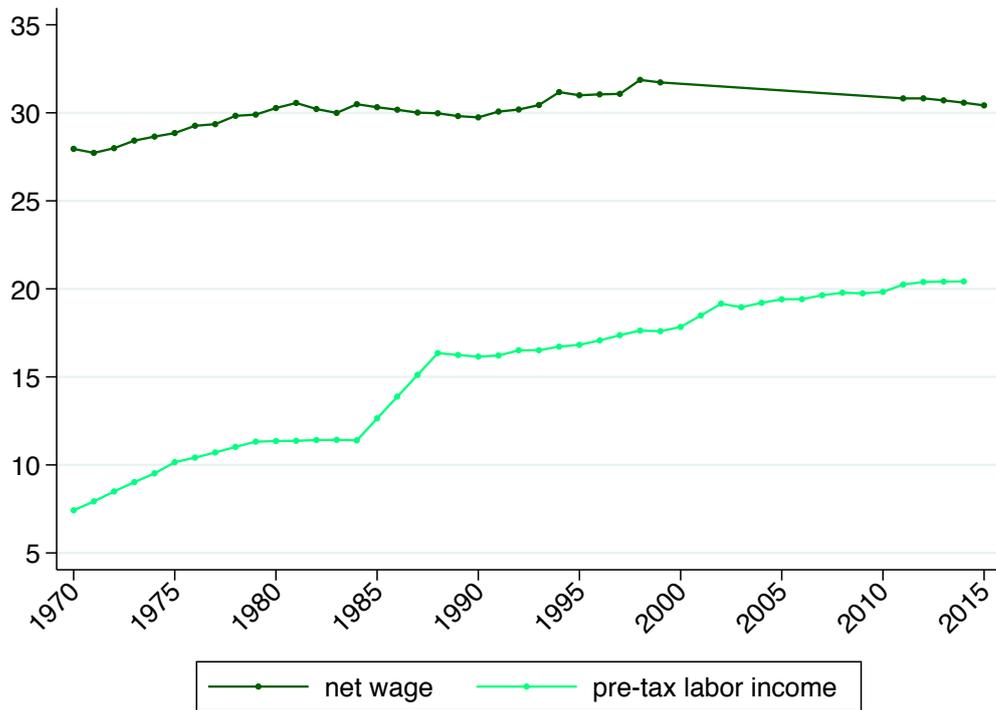
Graph 28b – Top 10%



Graph 28c – Middle 40%



Graph 28d – Bottom 50%



Notes: Graph 28 displays the ratios of the total labor income earned in each subgroup of the distribution on total labor income, on the one hand for wage earners in the private sector (net wages) on the other hand for the entire adult French population (pre-tax labor income).

Sources: DADS for net wage shares; series from Piketty, Garbinti and Goupille (2017) for pre-tax labor income shares.

The highest gap concerns the bottom 50%: in 1970, the bottom 50% pre-tax income share is smaller than 10%, whereas it is around 27% in net wage. This important difference is linked to the differences in terms of population composition between the two series: the pre-tax series cover all adults, including people with no labor income, which amounts to a significant share of the adult population in 1970. Besides people with no income, adding self-employed workers, whose distribution is more widely spread, tend to lower the bottom shares. Adding unemployed and retirees, whose income is on average lower, also pulls down the bottom share.

The gap reduced steadily from 1970 to 2000, with an important jump at the end of the years 1980, mostly because the pre-tax income share increased. This increase mainly resulted from the arrival of women on the labor market, which reduced the number of people with no labor income, and thus made the bottom 50% pre-tax income share increase much quicker than the fiscal income share. The gap seems to reach 10 percentage points in the years 2010.

Correspondingly, the top 10% and middle 40% shares are higher in pre-tax income, not because of the differences between the income concepts, but because the inclusion of individuals with null or low income makes the overall average income decrease more than the top 10% and middle 40% average incomes. Besides, the inclusion of self-employed workers probably pulls up the average income of the top of the distribution.

Both gaps decreased over the period: the decrease is small for the middle 40% (from 10 to 7 points), but more significant for the top 10% (from 10 to 4 points). Consequently the middle 40% pre-tax income share declined only slightly compared to the net wage income shares (except at the end of the years 1980, when the bottom 50% pre-tax income share increases strongly), and the top 10% pre-tax income share decreased more quickly than the fiscal income share.

This means that the pre-tax average income of the middle 40% wasn't much affected by the increase in the overall pre-tax average income triggered by the arrival of new workers (otherwise its share would have decreased much more). On the contrary the average pre-tax income of the top 10% didn't increase as much as the overall pre-tax average income, because it kept its own progression rhythm. During the years 2000, it was mostly the top 10% net wage share that increased, closing the gap even more. This could be related to the rise of very top wages in the private sector, which made the top 10% share less sensitive to the inclusion of other types of workers like self-employed workers.

The top 1% share is also higher in pre-tax income than in fiscal income, for the same reasons than for the top 10%. The gap remains relatively constant until the years 2000, but seems to have closed in 2011, as the net wage top 1% share has drastically increased, probably thanks to the rise of very high wages in the private sector.

Section 5. Directions for Future research

Further research should aim at building long-run series for the distribution of pre-tax labor income among the adult French population. There are three main challenges in this respect: correct for individuals missing from the wage data used in this work, include self-employed workers, retirees and unemployed, absent from the field of study before 1970, include adults with no labor income, also absent before 1970.

From 1919 to 1938: correct for the missing bottom 90%

Data on wages has been collected since the middle of the 19th century. The bulk of this data was produced by the Prud'hommes (labor courts). The data is sometimes very rich, with annual series of wages for a very detailed panel of jobs in a specific sector and location, for instance. But there was no systematic data collection, so some sectors are poorly represented, data wasn't collected on a regular basis, and figures are often local, (and given the spatial variability of remunerations at the time it is difficult to generalize local information).

Moreover, to construct a distribution from these figures, one needs data on the number of workers per type of job. Unfortunately, jobs headcount data is very poor in France until the middle of the 20th century: the main source is censuses, and these divide the labor force in only four categories: factory-workers (*ouvriers*), employees (*employés*), individual workers (*travailleurs isolés*), and employers (*chefs d'entreprise*).

Despite these limits, available data can inform about general trends in the bottom of the distribution. Indeed, the evolution of the remunerations of a relatively large number of jobs is reported on a very regular basis, sometimes yearly, during the interwar period (or sometimes before World War I). In particular, the section “Mouvements des salaires” (Wages evolutions) of the “Bulletin Mensuel de Statistique” produced by the *Statistique Générale de la France* (SGF, the predecessor of INSEE), displays this kind of series, with separated figures for Paris and the rest of France¹¹.

If we can identify at some point in time where these jobs are “located” in the wage distribution, following their remuneration over the years could suggest (approximately) how the corresponding percentiles of the distribution evolve, assuming that these jobs don't

¹¹ See Data Appendix 2 for an example.

“move” too much within the distribution. This assumption is of course simplistic, but it is probably valid for a certain number of jobs before World War II. Agricultural workers for instance, have always corresponded to the poorest wage earners, and during the first half of the century they approximately amounted to 10% of wage earners (though this proportion decreased). Thus, following the wage of agricultural workers could yield an index of the evolution of the average wage of the bottom 10% of the distribution.

The section “Mouvements des salaires” displays the evolution of the remunerations of a variety of working-class jobs (carpenters, plumbers, liners maid etc.) throughout the interwar period. However, data on employees (workers in banks or insurances for instance) is rare. But the remuneration of these is probably very close to their civil servant counterparts, on whom data is available.

There might also be data on self-employed workers. There are for sure regular figures on the number of employed workers per self-employed, which can be used as a first basic inequality indicator.

Between 1947 and 1970: from wage series to labor income series

There are several solutions to correct for the incomplete field of study of the DADS data. First, the DADS distribution could easily be merged with distributions for civil servants, on which data is available. Data should then be found on self-employed workers, retirees and unemployed.

Another possible solution relies on the comparison of DADS series with ERFS series (for the benchmark group of positive labor income earners). Indeed, ERFS series can be extended back until 1970: before 1996, tax returns are matched with census data and not *Enquête Emploi*, but census data is precise enough to allow the same distinctions (in terms of worker categories) than after 1996. As it has been seen in the previous section, there are important and understandable differences, in terms of income shares, between DADS and ERFS series. But these differences are not drastic, and observing them from 1970 to 2015 would enable to detect their evolution trends. Depending on the results of this comparison, it might be possible to compute ratios to apply to DADS series between 1947 and 1970, to upgrade them into labor income series (series for the entire population of adults with a positive labor income).

It should then be checked whether census data would enable to transform labor income series into series for the entire adult population (using census variables on the work status for instance).

From 1970 to 2015: towards a more precise decomposition of the labor income distribution

The richness of ERFs data is far from being exploited. Available information from the *Enquête Emploi* and from census data enables much more detailed decompositions of the labor income distribution than what was done in this work, by detailed job types for instance.

It would also be interesting to look at the sectorial composition of the various groups of the overall distribution (the share of self-employed workers, private and public wage earners in the top 10% of the distribution for instance).

Finally, the ERFs sectorial distributions could be upgraded to pre-tax income series using the procedure described in Piketty, Garbinti, Goupille (2016).

Gender inequalities

The information on gender available in the data has not been exploited in this study, but would enable to examine labor income gender inequalities. In ERFs, a gender indicator enables to do it in a straightforward way. In DADS data, for some years, there are separate tables per gender (organized the same way as the general tables). Before 1947 the information is much less straightforward, as there is no information on gender in the fiscal administration's summary statistics. However, the SGF data described in this section displays for certain years of the interwar period figures for specific "female's professions" ("professions féminines"), such as lacemakers, linenmaids, as well as general remuneration indexes for "female's professions" and "male's professions" (see Data Appendix 2 for an example). This data is very limited but it could provide evolution indexes.

Extending the series before World War I

Unfortunately, there is no fiscal data on wages before World War I, but there is probably more data on civil servants than what is presented in this work, and civil servants constituted an important part of the population of wage earners during the first half of the 20th century, all the more so before World War I. Thus, data on civil servants could be used to extend the series beyond World War I.

Section 6. Conclusion

In this study, I have analyzed in a systematic manner fiscal data on labor income in France using generalized Pareto interpolations, in order to examine the evolution of the entire distribution of labor income over the 20th century, and to decompose it between different types of workers for the recent period.

The results of this work confirm the periodization on which the literature agrees. After a reduction during World War I, wage inequalities increased until the beginning of the years 1930. High wages were then affected by the crisis of the years 1930, and most of all by the policies of the *Front Populaire*, and wage dispersion came back to its level of the early years 1920. Wage inequalities decreased even more during World War II, but post-war reconstruction of wage hierarchies was much quicker at the end of the years 1940. From 1950 to 1968, wage inequalities increased significantly, particularly driven by very high wages. This is probably a reason for the poor reputation of growth in the years 1960 in France, and for the social uproar of May 1968. From 1968 to 1983, wage dispersion was considerably reduced, mostly because of a much more generous policy towards the minimum wage. Between 1983 and 1999, the distribution stagnated, and the explosion of wage inequalities that took place in the Anglo-Saxon world didn't happen in France. But since 2000 inequalities seem to be on an increasing trend again, led by the rise of very high remunerations in the private sectors (for both wage earners and self-employed). So the question remains open whether France is following the same path than the United States a few decades late.

It is remarkable how constant income shares remained within the bottom 90%. Most of the fluctuations in overall wage inequalities were imputable to changes for the top 10% of the distribution. There has been very few “transfers” between the subgroups of the bottom 90%. The main noticeable trend is the increase of the income share of the very bottom of the distribution (below the 30th percentile), probably thanks to growing intolerance towards extremely low wages over the century (embodied in the creation of a minimum wage in 1950). This stability is all the more surprising given all the structural transformations experienced by the French labor market over the century. This stability confirms that most of

the decline in income inequality that took place in France over the 20th century is imputable to the evolutions of capital income inequalities.

The long-run evolution for civil servants is different: wage inequality in the public sector was drastically reduced around World War II, with very little evolution since then, making the wage hierarchy of civil servants much more compressed than in the private sector.

Looking at the decomposition of the labor income distribution between sectors for the recent period, it appears that the public sector is still much more equal than the private sector. Another clear demarcation is between wage earners and self-employed workers: the distribution is much more spread-out for the seconds, and the gap could be increasing in most recent years.

There is still much to be done to study the evolution of the distribution of labor income in France since the beginning of the 20th century. First, the data available for the period 1970 – 2015 enables more elaborate decompositions than the one presented here. Before 1970, the main limit of the data is its restriction to wage earners, and to top wage earners before 1947. But filling this gap looks possible and should be a topic for future research.

REFERENCES

- J-C. ASSELAIN, *Histoire Économique de la France du XVIII^e à nos jours*, tome 2: *Depuis 1918* (227p.), Éditions du Seuil (Collection “Points Histoire”).
- C. BAUDELLOT and A. LEBEAUPIN, 1979, “Les salaires de 1950 à 1975”, *Économie et Statistiques* n°113, p. 15-22.
- A. BAYET, 1997, “Deux siècles d’évolution des salaires en France”, INSEE, Document de travail de la Direction des Statistiques Démographiques et Sociales n°F9702 (29p.).
- A. BAYET and M. JULHÈS, 1996, “Séries longues sur les salaires”, *INSEE – Résultats* n°457 (série Emploi-Revenus n°105) (94p.).
- T. BLANCHET, J. FOURNIER, T. PIKETTY, 2017, “Generalized Pareto curves: Theory and Applications”, WID.World Working Paper 2017/03.
- B. GARBINTI, J. GOUPILLE-LEBRET and T. PIKETTY, 2016, “Accounting for Wealth Inequality Dynamics: Methods, Estimates and Simulations for France (1800 – 2014)” WID.World Working Paper 2016/05.
- B. GARBINTI, J. GOUPILLE-LEBRET and T. PIKETTY, 2017, “Income Inequality in France, 1900 – 2014: Evidence from Distributional National Accounts (DINA)”, WID.World Working Paper 2017/04.
- INSEE, 1951, *Recensement général des agents des services publics en 1950*.
- S. KUZNETS, 1953, *Shares of Upper Income Groups in Income and Savings*, National Bureau of Economic Research (707p.).
- S. KUZNETS, 1955, “Economic Growth and Inequality”, *American Economic Review* 45 – 1, March 1955, p. 1-28.
- V. PARETO, 1897, *Cours d’Économie Politique*, vol. 2. (réédition sous la direction de G. Busino, Librairie Droz, 1964).
- T. PIKETTY, 2001, *Les hauts revenus en France au XX^e siècle. Inégalités et redistributions, 1901-1998*, Grasset.
- T. PIKETTY, 2014, *Capital in the 21st century*, Harvard University Press.
- T. PIKETTY and E. SAEZ, 2003, “Income Inequality in the United States, 1913 – 1998”, *Quarterly Journal of Economics*.
- D. QUARRÉ, 1999, « Les salaires des agents de l’Etat en 1996 – Résultats détaillés », *INSEE-Résultats* n° 660 – 661, (série Emploi – Revenus n° 151 – 152).

R. RIVET, 1932, “Les statistiques des fonctionnaires en France et dans divers pays”, Bulletin de la Statistique Générale de la France.

V.TURQUAN, 1899, *Essai de recensement des employés et fonctionnaires de l'Etat, suivi d'une statistique des pensionnaires de l'Etat*, Paris, au secretariat de la Société d'économie sociale.

- Generalized Pareto Interpolation interface : <https://wid.world/gpinter/>

APPENDIXES

Appendix A – Summary Statistics on the *Impôt Cédulaire sur les Traitements et Salaires* (1919 – 1938)

Data source

The data on the distribution of wages between 1919 and 1938 comes from the monitoring by the French fiscal administration of the schedular tax on salary, wages, pensions and annuities (“*Impôt sur les traitements, salaires, pensions et rentes viagères*”).

This tax was created within the framework of the major French fiscal reform of 1914-1917, along with the “*Impôt général sur le revenu*” (general income tax). It was an individual tax applying to all wages above a certain threshold that increased from 6 000 francs in 1919 to 10 000 francs in 1938. It was first implemented in 1917, and from then it became mandatory for all employers to declare to the fiscal administration the precise annual amount of the wages paid to each of their employees.

The tax administration used these employer tax returns to produce annual national tables on taxable wages and pensions. From 1919 to 1921, there were separate tables for wages on the one hand, and pensions and annuities on the other hand, so between 1919 and 1921 the tables concern wage earners only. From 1922 until 1938, wages and pensions are treated in a single table, and thus cannot be distinguished. However, according to Piketty (2001), pensions and annuities represent a relatively small portion of the overall tax base, so the bias they introduce can be overlooked. From 1919 to 1930, tables are in net wage, but from 1931 to 1938 they are expressed in taxable wage. From 1934 to 1938 the difference between net and taxable wages is negligible, as there was no major tax deductions, but from 1931 to 1933, family deductions introduce a significant difference between net and taxable wages.

The tables are organized by earnings brackets: for each bracket, the total number of workers whose wage (or pension) falls into that bracket, and the total amount of wages or pensions received by these workers, are indicated. The base unit is the individual (not the household), since the *impôt cédulaire* is individual.

The tables include all the individuals who actually paid the schedular tax on wages, whatever their occupation and sector is (public or private), but only them. This has two important consequences.

First it implies that they give no information on non-taxable workers (that is to say the majority of wage earners during the interwar period). Overall, assuming a total population of approximately 12 million wage earners over the interwar period, non-taxable workers correspond to between 63 and 83 % of wage earners, depending on the year. Thus these tables provide information only on the top of the wage distribution. I assume a total population of 12 million wage earners following Piketty (2001). This simplifying assumption is due to the fact that the number of wage earners indicated in censuses fluctuated around 12 millions during the interwar period. Otherwise, there is no satisfying series on the number of wage earners during this period, in particular because of biased introduced by different methods of unemployment measurement.

Second, tables are truncated from 1922 to 1933, because individuals could then benefit from tax deductions if they were married or had children or dependents, (called “family deductions” in this study, for simplicity). Those who became non-taxable thanks to such deductions are thus absent from the tables, even if their wage was above the tax threshold. This results in underestimating the number of people at the top 10% or 20% of the wage distribution between 1922 and 1933.

Correction for truncated tables because of family deductions

From 1922 to 1930, since tables are expressed in net wage, it is possible to correct for missing individuals in a systematic way, using the information on the total amount of the various types of tax deductions available in the original tables. Four variants of the same general method are used, depending on the deductions rules and on the number of thresholds displayed in the tables: one for 1922, one from 1923 to 1925, one in 1926 and 1927, and one from 1928 to 1930.

The basic variant is the one used in 1926 and 1927. The tax threshold was then of 7000 francs, married couples had a tax deduction of 3000 francs, each child gave right to an additional deduction of 3000 francs, and each dependent gave right to an additional deduction of 2000 francs. For families where both parents were taxed, each parent was taxed separately, but deductions were only granted to the man: married women were considered as singles with

no children. Original tables display thresholds from 7000 to 50 000 Francs. They also indicate for each earnings bracket and type of deduction (marriage, children, and dependents) the total amount of deductions granted in that bracket (see in Data Appendix 1 for an example of an original table).

In 1926 and 1927, married individuals with seven children and four dependents had a tax threshold of 39 000 francs. Consequently, in the (40 000 - 50 000 francs) bracket, only individuals with more than seven children and four dependents were absent from the table. These families were very few and can be neglected. Thus, it can be assumed that no one is missing from the two last brackets of these tables (40 000 – 50 000) and (50 000 and above), and the information on family deductions in these brackets can be exploited to get a (very blur) picture of family structures.

For each bracket, dividing each “amount of deductions” figure by the corresponding deduction (3000 or 2000) yields an approximate number of granted deductions, for each type: the “Amount of deductions for family situations” yields a number of married couples, the “Amount of deductions for family charges at 3000 francs” yields a total number of children, and the “Amount of deductions for family charges at 2000 francs” yields a total number of dependents. However, the two last figures correspond to a number of children or dependents, not a number of families, so they don’t give any direct information on the number of taxpayers who benefited from children (or dependent) deductions. Besides, if we assume that most taxpayers with underage children were married, the marriage deduction category includes the children deduction category: the majority of taxpayers who were granted children deductions were also granted marriage deduction (the reciprocal is not true). Consequently, if a wage earner is missing from the table because of children deductions, and is thus missing from the children deduction category, there is a strong probability that he is also missing from the marriage deduction category. This no-single parent assumption is reductive, but it seems relatively plausible to assume that most declared children had married parents during the interwar period (however it overlooks the case of widows).

I use the average number of children per family from the 1926 census to build mutually exclusive categories. The 1926 census gives the number of families per family size (from no child to 7 children). I use these figures to build an indicator of the average family size conditional on having at least one child. Dividing the number of children deductions by this average family size yields a number of families, and thus a number of taxpayers granted at least one child deduction. Then I subtract this number of taxpayers with children to the

number of marriage deductions, which is supposed to yield the number of married (male) taxpayers with no children, (under the assumption that only married individuals declare children). This assumption could be refined in future research, by subtracting only a fraction of the number of taxpayers with children, to take into account single parents. Thus, for highest brackets in which no one is missing, I obtain the number of individuals in two mutually exclusive categories: taxpayers with children (assumed to be married), married taxpayers with no children.

Wage earners can also be absent from tables because of a combination of deductions for marriage, children, and dependents, thus information on the probability to have one or more dependent deductions is also needed. To compute this probability, I again use the assumption that no one should be missing in brackets above 40 000 francs. Thus, the ratio of the number of deductions for dependents on the total number of tax payers in the bracket yields the average number of dependents per individuals in the 40 000 – 50 000 francs bracket, called D.

A dependent can be either an older parent or a child above 18 years old. The probability of having an older parent at one's charge is fixed to 0.05 in the 40 000 – 50 000 francs bracket. This probability is roughly estimated using data from the lowest bracket (7000 – 10 000 francs): in this bracket, only single tax payers with no more than one dependent at charge are present, (all married people, and even single parent with at least one child are exempted from the wage tax in this bracket). Thus we know that the number of deductions for dependents here corresponds to the number of individuals who have only one parent at their charge (and whose wage is above 9000 francs). Assuming linear repartition of individuals within the bracket, we divide the number of deductions for dependents by $(N/3)$, N being the total number of people in the bracket. This can be considered to yield the probability of having a parent at charge (for singles). It is roughly equal to 0.03, and I assume it increases with income (because of the positive correlation between age and income on the one hand, and the plausible positive correlation between age and the probability to have an older parent at charge). However I assume this probability is the same for all individuals of the same bracket (whether they are single, married, or have children). Thus, I fix it (P_{parent}) at 0.05 for the 40 000 – 50 000 bracket.

I also assume that dependent older parents are “independent events”, so that the probability of having two older parents as dependents is equal to 0.05^2 , and so on. I also assume that having an adult child at charge is independent from having an older parent at charge. Then, I assume that the ratio r_n between the probability of having n adult children at charge (P_{nch}) and

the probability of having one adult child at charge (P_{1ch}), is equal to the ratio of the probability of having n children and the probability of having one child, which can be computed with the 1926 census data on family size. I suppose the proportion of people having four dependents or more is negligible and fix the probability of these events to 0. Given these assumptions, I can compute the probability of having n dependents (P_{ndep}) for n smaller than 4, as a function of the probability of having one dependent adult child P_{1ch} :

$$\begin{aligned}
 - P_{1dep} &= P_{1parent} + P_{1ch} - P_{1parent} * P_{1ch} \\
 - P_{2dep} &= (P_{1parent})^2 + P_{1ch} * P_{1parent} + r_2 * P_{1ch} \\
 - P_{3dep} &= (P_{1parent})^3 + P_{1ch} * (P_{1parent})^2 + r_2 * P_{1ch} * P_{1parent} + r_3 * P_{1ch}
 \end{aligned}$$

The only unknown parameter is P_{1ch} , which I can derive by solving the following equation:

$$D = 1 * P_{1dep} + 2 * P_{2dep} + 3 * P_{3dep}$$

With D the average number of dependents per taxpayer, computed earlier. This equation can be rewritten as:

$$P_{1ch} = \frac{A - P_{1parent} - 2 P_{1parent}^2 - 3P_{1parent}^3}{1 + P_{1parent} + 2 r_2 + 3P_{1parent}^2 + 3r_2 P_{1parent} + 3r_3}$$

Using P_{1ch} , the probabilities of having one, two, or three dependents can then be derived.

Using all this information from the higher brackets, I can then correct for missing individuals in the lower brackets.

For the (30 000 – 40 000) and the (20 000 – 30 000) brackets, I create the same categories than for higher brackets (married individuals without children, married individuals with children), except that I adapt the average family size by which I divide the number of deductions for children, to account for potentially missing individuals.

To do so, I assume that people with n children are linearly spread in each income bracket. Thus I discount the number of families with n children from the census (N_n) by their probability to be in the missing group. That probability is equal to the sum, for x going from 0 to 3, of the products between their chance to have x dependents and the portion of the income bracket for which they would become not taxable with x dependents. In the (30 000 – 40 000) bracket for instance, we know that individuals with 5 children, 3 dependents, and earnings below 31 000 francs are missing, as well as individuals with 6 children, 2 dependents and an income below 32 000 francs, and so on. Consequently, the average family size is

$$F = F_1/F_2$$

$$\begin{aligned}
\text{With } F_1 &= (N_1 + 2 * N_2 + 3 * N_3 + 4 * N_4 + 5 * N_5 * (1 - P_{3dep} * 0.1) \\
&+ 6 * N_6 * (1 - P_{2dep} * 0.2 - P_{3dep} * 0.4) \\
&+ 7 * N_7 * (1 - P_{0dep} * 0.1 - P_{1dep} * 0.3 - P_{2dep} * 0.5 - P_{3dep} * 0.7)) \\
\text{And } F_2 &= N_1 + N_2 + N_3 + N_4 + N_5 * (1 - P_{3dep} * 0.1) \\
&+ N_6 * (1 - P_{2dep} * 0.2 - P_{3dep} * 0.4 - P_{3dep} * 0.6) \\
&+ N_7 * (1 - P_{0dep} * 0.1 - P_{1dep} * 0.3 - P_{2dep} * 0.5 - P_{3dep} * 0.7))
\end{aligned}$$

It can be considered that all married individuals without children with earnings above 20 000 francs are taxable (they would need to have at least four dependents to become non-taxable). Thus the number of married individuals without children matches reality in the (20 000 – 30 000) and the (30 000 – 40 000) brackets. Thus I can apply to them the ratio between individuals with children and married individuals with no children computed in the (40 000 – 50 000) bracket, to obtain a new number of individuals with children for the (20 000 – 30 000) and (30 000 – 40 000) brackets. To correct for total earnings, I assume that the individuals I added have the average earnings of the bracket (so the average earnings don't change), except in a few cases when I know that the people added necessarily earn less than a certain threshold. In these cases I assign to them earnings equal to the unweighted average of their reduced bracket.

For the (10 000 – 20 000) bracket, I proceed the same way, except that even the category of married people without children is truncated. I correct for missing individuals in this category by assuming that people are evenly distributed over the whole bracket. Married individuals with one dependent are not taxable under 12 000, the threshold is 14 000 for those with two dependents and 16 000 for those with three dependents, so I assume that the number of observed individuals m is equal to $((P_{0dep} + 0.8 * P_{1dep} + 0.6 * P_{2dep} + 0.4 * P_{3dep}) * n)$, with n the “true” number of individuals that should be counted in that bracket. Thus I correct m accordingly and apply the ratio between married individuals without children and married individuals with children obtained from higher brackets.

It is impossible to correct the (7000 – 10 000) bracket with this method, because no married people, with or without children, are taxable in this income bracket. The only deduction that can still apply for individuals in that bracket is for dependents, and we cannot derive a corresponding ratio from higher brackets. Indeed in higher brackets, this deduction

category captures deductions for all types of individuals (married, with children and so on), whereas in the lowest bracket it only captures deductions for singles with no children, so they are not really comparable.

From 1923 to 1925, the same legislation applies, but the highest bracket of the tables is 20000 francs and more. This makes it more difficult to properly discount the average family size used to divide the number of deductions for children: since there is no finite upper boundary, I cannot directly discount the average by an assumed portion of missing individuals. Instead, I gather the four upper brackets of the year 1926 (20 000 – 30 000, 30 000 – 40 000, 40 000 – 50 000, 50 000 and more) in a single bracket (20 000 and more) and derive the ratio of the “true” number of individuals that I computed and the number of observed individuals in 1926. I apply this ratio to the number of observed people in the 20 000 and more bracket of 1923, 1924 and 1925. To correct total income, I proceed the same way and apply the ratio between “true” total income and observed total income for the four upper brackets of 1926.

The lower brackets (15 000 – 20 000 and 10 000 – 15 000) are corrected the same way as in 1926 and 1927, but I cannot derive proper category ratios in the upper bracket, because categories are truncated in the 20 000 and more bracket, and I only corrected the total headcount and income. Thus, I use category ratios from 1926.

There is the same issue for 1922 as the highest bracket is 20 000 francs and more. Moreover, the legislation is not the same: the deduction for children is 2000 francs per child, and the deduction for dependents is 1500 francs per dependent. Consequently, I cannot apply the ratio between “true” and observed figures from 1926 (as not the same number of people are not taxable).

Approximately no tax payer with earnings above 30 000 francs is missing given the new deduction rules, thus we only have to take into account missing people whose income is between 20 000 and 30 000 francs in the highest bracket. Consequently, I compute a “bracket weight” using 1926 figures, that is to say ratios of the observed headcount of the 20 000 – 30 000 bracket on the total observed headcounts of all wage-earners above 20 000 francs. I then compute the average family size by discounting the discount itself (that accounts for missing individual) with the bracket weight. With this corrected family size, I build the usual categories and use the ratios from 1926 to correct for missing individuals with children.

Inferior brackets are corrected with the basic method, using ratios from 1926.

From 1928 to 1933, the tax threshold is 10 000 francs, the deduction for marriage is 3000 francs, the deduction for children is 3000 francs for the two first children and 4000 francs from the third child. The deduction for dependents is 2000 francs. From 1928 to 1930, the tables' highest threshold is 50 000 francs, and they display an additional category: deduction for children of 4000 francs, which enables to get the number of children that are at least the 3rd in their family, and so the number of families with more than two children. Thus, I apply the exact same method than for previous year using this additional category.

From 1931 to 1933, the tables are expressed in terms of taxable earnings. This means that an individual counted in a certain bracket might actually be in the superior bracket in terms of net wage. It is impossible to come back to net wage tables because we cannot derive the proportion of a bracket whose net wage is in the higher bracket.

Consequently, I only corrected the final results of the Pareto interpolation, in very *ad hoc* way, to ensure that they do not disturb longer trends. I followed in that the method used by Piketty (2001), which consists in adding to thresholds and average income levels amounts that approximately match the average deduction observed in the original table. This result in adding:

- 2250 francs to the average income of p90p95
- 2500 francs to P90
- 3000 francs to the average income of p95p99 and to P95
- 4000 francs to the average income of p99p99.5 and to P99
- 5000 francs to the average income of p99.5p99.9 and to P99.5
- 6000 francs to the average income of p99.9p99.99, of p99.99p100, and to P99.9 and P99.99.

Appendix B – *Déclarations Annuelles de Données Sociales (1947 – 2015)*

Data source:

The tax administration stopped to produce national tables in 1938. But the obligation for employers to declare annual individual wages paid to any employee remained, and INSEE started using employer tax returns in 1947 to produce similar summary statistics.

Like those of the interwar period, INSEE's tables are expressed in terms of net individual wage, that is to say any wage paid to the employee, net of contributions to the pension system, to unemployment insurance and to social insurances. Wages include the "Contribution sociale généralisée" (CSG) until 1996, but from 1997 onwards they are net of CSG and of "Contribution au Remboursement de la Dette Sociale" (CRDS)¹². They also include taxable indemnities.

The tables display, for a certain number of earnings brackets covering the whole distribution, the number of individuals whose earnings fall into the given bracket. From 1947 to 1952 (except in 1950), and from 1993 to 1999, the total amounts of all the wages paid (per bracket) are also indicated, like in the interwar tables. However from 1956 to 1992, only headcounts and earnings thresholds are given.

Unlike the tax administration, INSEE's goal was not to monitor a tax, but to produce statistics on wages. Consequently, all wage earners were taken into account, even if they were not taxable. Thus, tables cover the entire distribution, which is a major strength compared to previous tables and to similar tables of the same period for income and wealth.

However, a few categories of wage earners are absent from these tables, mostly because the declarations filled by employers were different and didn't enable the same kind of statistical exploitation. Public workers (civil servants and contract workers alike) are excluded. But workers from State agencies that have commercial or industrial activities (publicly owned companies, public hospitals for instance), as well as those from local communities, are included since 1992. Domestic workers are excluded throughout the whole period, agricultural workers are included since 2002, but the tables I use after 2002 exclude them. Finally, between 1952 and 1963, the tables include only permanent workers. Part-time workers are progressively included in the field of study since the end of the 20th century, but

¹² Contribution Sociale Généralisée and Contribution au Remboursement de la Dette Sociale are two taxes respectively created in 1991 and 1996, deducted at source.

the tables I use exclude them until 2009. For the 2009-2015 period, the tables I use include part-time workers' wages expressed in "full-time equivalent" ("équivalent temps-plein"), that is to say corresponding to what the workers would earn if their job was full-time. Interns and apprentices are included in the field of study, but from 2012 to 2015 I use tables excluding them.

From 1947 to 1996, I use the exact same data as Piketty (2001) (the references of the sources are available in Data Appendix 3).

From 1997 to 1999 I use the same sources than the ones used for previous years, except that tables are expressed in hourly wages. I kept hourly amounts to compute all the relative results (ratios of thresholds on average wage, income shares). To derive average wages, I converted all the values of the input data in annual values by multiplying the average hourly wage of each bracket by the average number of hours per bracket, (computed by dividing the total number of hours declared by the total number of people). From 2004 to 2015, wages are expressed in monthly terms, so I multiplied all necessary values by 12 for absolute results, and used monthly values for relative results.

INSEE's tables's thresholds are most of the time not high enough to estimate properly the very top of the distribution. The highest threshold rarely corresponds to a percentile higher than the 99.5th percentile, and at the end of the years 1960 and of the years 1980, the 99th percentile isn't even covered.

Moreover, INSEE doesn't seem to publish similar tables to those described above since 2000. For the years 2000, there are distributional tables (with thresholds and headcounts only), produced by INSEE using DADS. Between 2000 and 2010, information is displayed at the decile level only, and since 2011 it is available at the centile level. Consequently, I only used tables from 2011 to 2015, because tables from the years 2000 were not detailed enough to yield reliable estimations.

Post – estimation corrections:

INSEE didn't produce any national tables for the years 1948, 1949, 1953, 1958, 1981, 1983 and 1990. There were tables in 1954 and 1955, but their superior bracket was not high enough to produce satisfying estimations of average income levels above the 90th percentile, so I didn't use them.

For the year 1954, I followed Piketty's methodology (2001). I used the estimations of the 10th, the 50th, and the 90th percentiles from Bayet and Julhès (1996), to which I applied a ratio to upgrade them to estimations for all workers (including non-permanent workers), like for other years between 1953 and 1962. I then derived the following ratios for 1952 and 1956: $P99/P90$, $p90p100/P90$, $p99p100/P90$, $p50p90/P50$, $p0p50/P50$, and averaged them to get values for 1954. I applied the 1954 ratios to the corrected percentiles of Bayet and Julhès to obtain values of the average wage of the top 1%, the top 10 %, the middle 40% and the bottom 50% of the wage distribution in 1954, as well as an estimation of the 99th percentile.

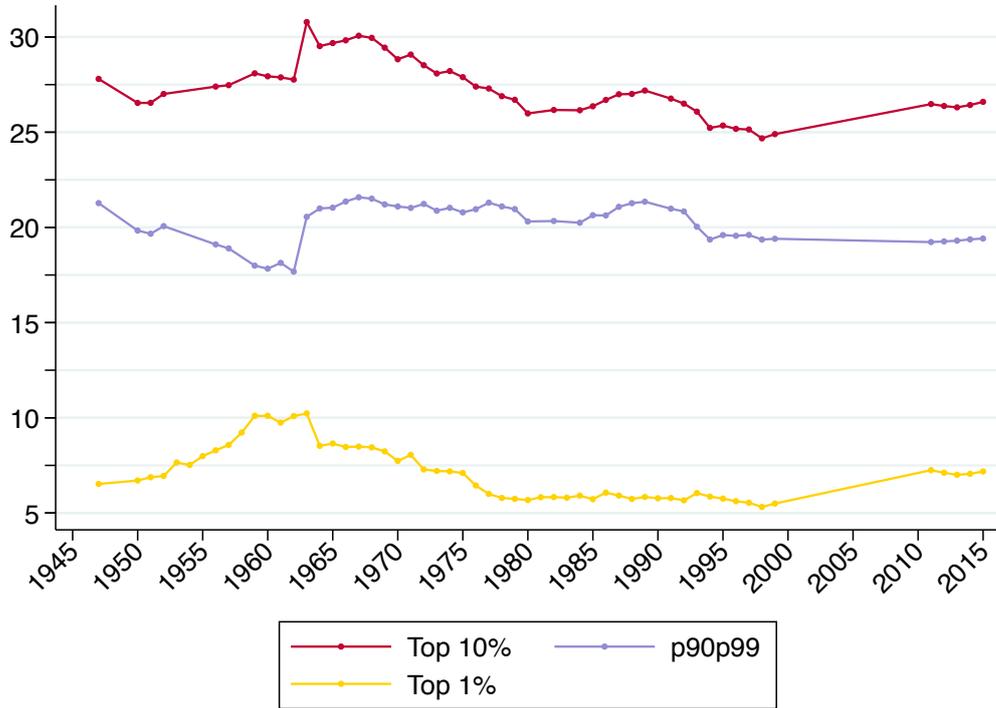
For the rest of the missing years, I computed thresholds and earnings levels by linear interpolation.

Corrections for non-permanent workers (1953 – 1962):

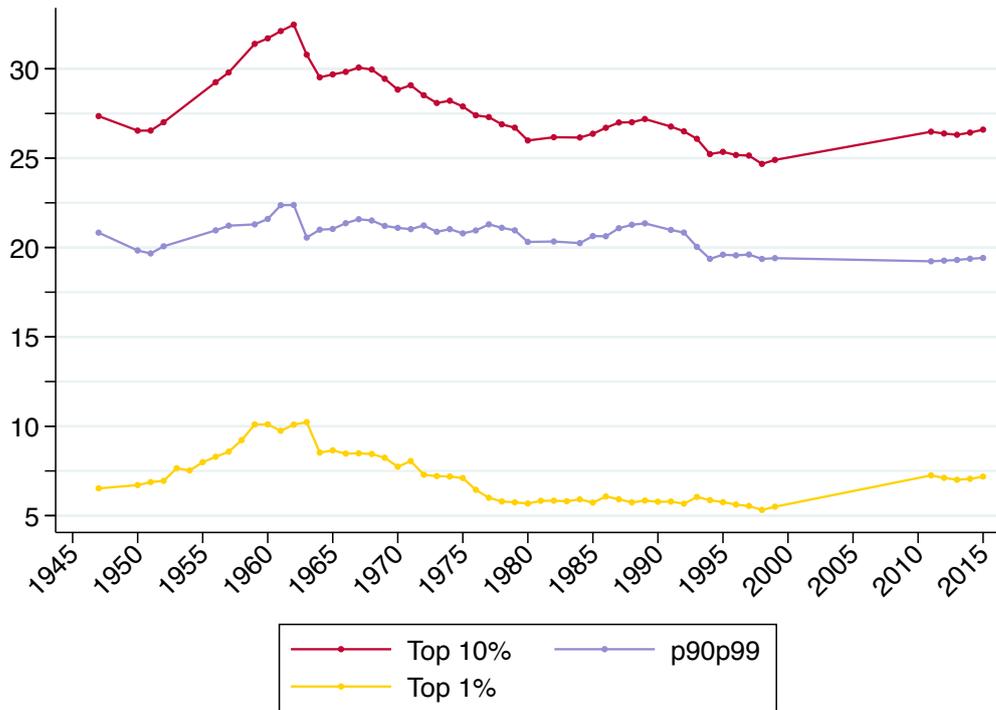
The corrections applied to wage levels and thresholds between 1953 and 1962 to account for the absence of non-permanent workers, described in Section 3, yield new wage shares for each subgroup of the distribution that do not necessarily sum up to 1, because the correction coefficients are derived independently for each subgroup. To have consistent shares, I recomputed the top 10% share by subtracting all new shares to 1. I chose to adapt the top 10% to the rest because the inclusion of non-permanent workers is likely to have a more important impact on bottom groups. I present here the two versions of the evolution of the top 10% share and within: the first (AB – 1) with the top 10% corrected to match other shares, the second (AB – 2) with the top 10% corrected using the ratio method, not adapted to other shares.

The real trends for the Top 10% and p90p99 probably lies between the two graphs. In particular, the decrease of the share of p90p99 on Graph AB - 1 is probably overestimated because the top 10% share is shrunk to match other shares whereas the top 1% share is upgraded like the rest.

Graph AB – 1 – Wage Shares within the top 10%, correction 1 (1947 – 2015)



Graph AB – 2 – Wage Shares within the top 10%, correction 2 (1947 – 1967)



Appendix C - Average Wage

For average wages, we used for the vast majority of years the values derived by Piketty (2001) the following way.

Between 1951 and 1992, Piketty used the series of annual net wage for full-time workers published in Bayet and Julhès (1996), constructed using DADS data. He just discounted Bayet and Julhès's value for 1950 by 14% to account for the fact that very low wages were excluded from DADS tables in 1950.

He then relied on series of annual net factory worker wages from 1900 to 1996, constructed by Bayet (1997), to compute the ratio of the average wage of all wage earners on the average wage of factory workers, from 1950 to 1996. He used wage declarations to derive a value for 1947, and assumed that the ratio increased linearly from 1 in 1900 to its 1947 value (1.16). He then applied this ratio to Bayet's series on factory worker wages to obtain average wage values for the years before 1947.

I used Piketty's series except for years for which DADS tables indicated average earnings levels for each bracket for the whole distribution, that is to say 1947, 1951, 1952, and from 1993 to 1999. For these years, I computed the average wage from the per-bracket averages.

Between 2000 and 2015, I used the average wages given on the distributional tables released by INSEE.

Appendix D - Civil Servants

The data I used for civil servants comes from different sources but has the same format than the tables used for other wage earners. The tables for 1896 and 1931 respectively come from the seminal researches of Victor Turquan (1899) and Raymond Rivet (1932) on civil servants. They both used statistics from budget reports or public finance reforms projects. The distribution of wages for civil servants is known to be very similar in 1896 and 1913¹³, so for simplicity I consider that the results of 1896 are also valid for 1913, and replace 1896 by 1913 in this study. The table for 1951 comes from the census of public workers conducted by INSEE in 1950. The data for 1996 comes from INSEE's survey "Les salaires des agents de l'Etat en 1996", for which various sources were compiled, such as individual files from various ministries, complementary questionnaires, DADS and census data. The table covers workers in State services *stricto sensu*, but also teachers in private schools that have a contract with the State, and workers in public institutions or public operators (there were two public operators in 1996: La Poste and France-Télécom). They do not include workers from public companies (such as transport companies for instance (SNCF, RATP)).

All the tables have the same format: for a large number of income brackets, they indicate the number of public workers whose wage falls into that bracket. The brackets cover the whole distribution, and capture rather well the top of the distribution. However there is no information on the average earnings for each bracket, only threshold information.

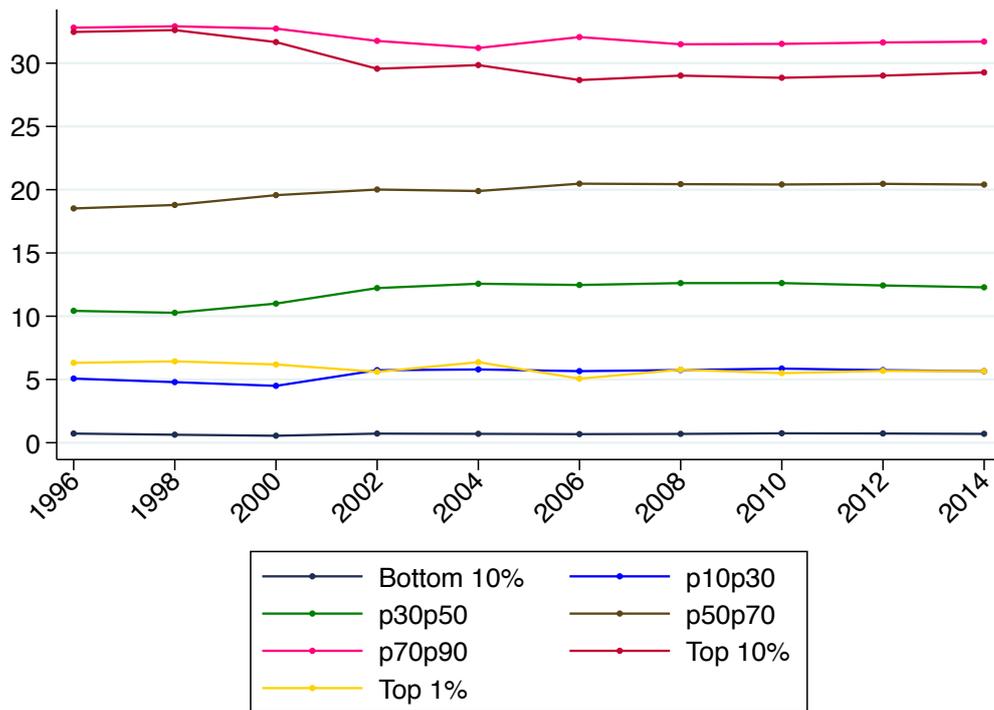
For 1896 and 1996, average earnings were indicated in the source. For 1931 and 1951, average earnings were calculated assuming linear repartition of income within each bracket (to derive total earnings per bracket). For the top bracket, for which there is no upper boundary, I followed Piketty's method (2001) and multiplied the lower threshold by 1.4 to compute the average earnings.

The table for 1931 has a 0 threshold for the lowest bracket, which is inconsistent with other years for which the threshold is always superior to 0. Consequently, I replaced the threshold by 300.

¹³ Piketty (2001)

Appendix E – Distribution of Pensions and Unemployment Benefits (1996 – 2014)

**Graph AE – 1 – Income Shares for Pensions and Unemployment Benefits
(1996 – 2014)**
(in percentage of total income)



Notes: Graph AE-1 displays the ratios of the total of pensions and unemployment benefits earned in each subgroup of the distribution (of pensions and benefits) on total pensions and unemployment benefits. See Graph 2 for the definitions of subgroups.

Sources: ERFS