TOP INCOME SHARES IN THE LONG RUN: AN OVERVIEW

Thomas Piketty
ENS-EHESS, Paris-Jourdan

Abstract
This paper offers an overview of what we have learned from a collective research project on income distribution in the long run. Using historical income tax statistics and a common methodology, we have constructed annual top income shares series (often broken down by income source) for over 20 countries covering most of the 20th century. One important conclusion is that the decline in income inequality that took place during the first half of the 20th century was mostly accidental, and does not seem to have much to do with a Kuznets-type process. Top capital incomes were hit by major shocks during the 1914–1945 period, and were never able to fully recover from these shocks, probably because of the dynamic impact of progressive income and estate taxation. Our database also allows us to readdress the cross-country analysis of the interplay between inequality and growth with better prospects than with standard databases. (JEL: D31)

1. Introduction

This paper presents some of the key findings and perspectives emerging from a collective research project on the dynamics of income and wealth distribution. The primary objective of this project was to construct a high-quality, long-run, international database on income and wealth distribution using historical tax statistics. The resulting database now includes annual series covering most of the 20th century for over 20 (mostly Western) countries. It seems about the right time to take stock and see what we have learned in this project and where we are going.[1]

    The primary motivation for this project was a general dissatisfaction with existing income distribution databases. Existing international databases on inequality are not high-quality (they display little homogeneity over time or across

---

E-mail address: piketty@ens.fr or piketty@ehess.fr

1. The present paper focuses mostly on the French findings. Some of the other countries are covered in Atkinson (2005, this issue), Dell (2005, this issue), and Saez (2005, this issue), and the references therein. See also the volume edited by Atkinson and Piketty (2005), which includes a full set of country chapters and series. The countries covered in the database so far include: France, the U.S., the U.K., Germany, Canada, Australia, New Zealand, Ireland, Switzerland, the Netherlands, Belgium, Spain, Italy, Norway, Sweden, Finland, Denmark, Japan, India, Argentina, and Brazil.

countries),[2] they are not long-run (they typically cover only a couple of isolated years per country, generally restricted to the post-1970 or post-1980 period), and they almost never offer any decomposition of income inequality into a labor income and a capital income component. This latter feature of existing data sets is unfortunate, because the economic mechanisms at work can be pretty different for the distribution of labor income (demand and supply of skills, labor market institutions, etc.) and the distribution of capital income (capital accumulation, credit constraints, estate taxation, etc.), so that it is fairly heroic to test for any of these mechanisms using such data. The fact that existing databases are not long-run is also most unfortunate, because structural changes in income and wealth distributions are relatively slow and very often span over several decades. In order to properly understand such changes, one needs to be able to put them into broader historical perspective.

Although our database also suffers from strong limitations (in particular, our long-run series are generally confined to top income and wealth shares and contain little information about bottom segments of the distribution), the fact that it is fully homogenous across countries, annual, long-run, and broken down by income source offers a unique opportunity to better understand the dynamics of income and wealth distribution and the interplay between inequality and growth.

The rest of this paper is organized as follows. In Section 2, we present the basic methodology used to construct the database and some of the main descriptive findings. Section 3 argues that our database can potentially be used to re-address the cross-country analysis of the interplay between inequality and growth, with better hopes of success than the previous literature. We then discuss some of the prospects for extending our database beyond the 20th century (Section 4) and beyond OECD countries (Section 5).

2. Constructing a New Data Base: Basic Methodology and Results

Household income surveys are a relatively recent venture: they virtually did not exist prior to 1950, and in most countries they were not available in an homogenous, machine- readable format until the 1970s–1980s. The only data source that is consistently available on a long-run basis is tax data. Progressive income tax systems were set up in most Western countries at the beginning of the 20th century (1907 in the U.K., 1913 in the U.S., 1914 in France, etc.), and in all countries with an income tax system the tax administration started compiling and publishing

---

2. See e.g., the Atkinson-Brandolini (2001) criticism of the World Bank (Deininger–Squire) secondary database. The D–S database is "secondary" in the sense that it is based on the collection of inequality measures computed by others using various income data sets and methodologies for different countries and time periods. In contrast, our inequality measures were computed by ourselves using the same primary data sources and methodology for all countries and time periods.

tabulations based on the exhaustive set of income tax returns.[3] These tabulations   **Fn 3**
generally report for a large number of income brackets the corresponding num-
ber of taxpayers, as well as their total income and tax liability. They are usually
broken down by income source: capital income, wage income, business income,
etc. One can then use standard Pareto extrapolation techniques to compute top
fractiles thresholds and average incomes using such data.

One major limitation of tax data is that the income of individuals not subject
to the tax is excluded from the data. Prior to World War II, the proportion of
individuals subject to progressive income taxation hardly exceeded 10–15% in
most countries, so that one can only compute top decile income series (and above)
over the entire period. In order to construct top fractiles income shares series from
top fractiles income series, one needs a total income denominator, which can be
computed using aggregate income sources (national accounts and their ancestors).

The first economist to use these data sources and methodology in a system-
atic way was Kuznets (1953). He exploited U.S. income tax tabulations covering
the 1913–1948 period and computed corresponding top decile and top percentile
income shares series. These were the first income distribution series ever produced
(income distribution had been at the center of speculative economic thought at
least since the time of Ricardo and Marx, but little data was available). Unsur-
prisingly, these series had a major impact on economic thinking, especially after
Kuznets proposed his famous "Kuznets curve" theory in order to account for the
1913–1948 decline in income inequality that he witnessed for the U.S.[4]   **Fn 4**

In a sense, all what we are doing is to extend and generalize what Kuznets
did in the early 1950s—except that we now have 50 more years of data and
over 20 countries rather than one. In addition, note that Kuznets had access to a
fairly limited data-processing technology, which probably explains why he did
not use all available data as systematically as possible. In particular, Kuznets did
not fully use the tabulations broken down by income source, and his top income
shares series are only defined for total income (for instance, he did not compute
separate series for wage income or capital income).

The fact that we have 50 more years of data, over 20 countries and series
broken down by income source led us to adopt a fairly different perspective as to
why income inequality dropped in Western countries during the first half of the
20th century. First, as one can see on Figure 1, where we plot the basic series for   **Figure 1**
the French case, the decline in top income shares witnessed by Kuznets for the
U.S. also took place in France, but it came to an end right after World War II.
The secular decline in income inequality took place during a very particular and

---

3. Full details about the administrative publications where the raw tabulations were originally
published are given for each country in Atkinson and Piketty (2005).

4. See Kuznets (1955). According to this theory, income inequality should follow an inverse-U
shape along the development process, first rising with industrialization and then declining, as more
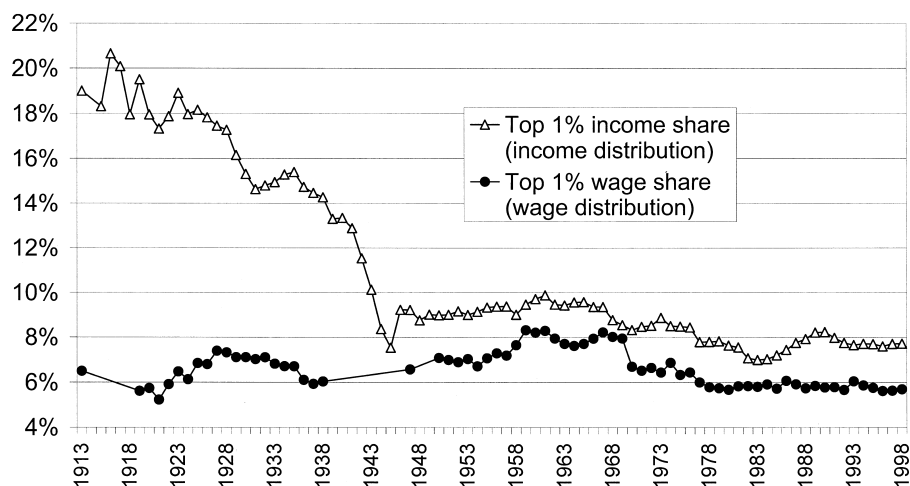and more workers join the high-productivity sectors of the economy.

FIGURE 1. The fall of top capital incomes in France, 1913–1998. Source: Author's computations using income tax returns (see Piketty 2001, 2003).

politically chaotic period, namely during the 1914–1945 period (and especially during both World Wars and the early 1930s). This raises serious doubts about a gradual, Kuznets-type explanation. If the decline in income inequality was due to a continuous reallocation process between from a low-productivity to a high-productivity sector (say, from rural to urban sector, as in Kuznets' original model), then it is hard to understand why the timing of the fall should be so peculiar.

Next, and most importantly, one can see from Figure 1 that the 1914–1945 drop in top income shares is entirely due to the fall of top capital incomes: top wage shares actually did not decline at all. One gets the same picture by using other inequality measures, e.g., by looking at the top decile share rather than the top percentile share. In particular, the striking fact that the wage distribution in a country like France has been extremely stable in the long run during the 20th century appears to be very robust, irrespective of how one measures wage inequality (for instance, the 90–10 ratio—and not only top wage shares— has also remained stable in the long run).[5] Labor reallocation of the kind described by Kuznets did take place (the bottom 30% of the French wage distribution was made up almost exclusively of rural workers at the beginning of the 20th century, and rural workers have virtually disappeared by the end of the 20th century), but this did not lead to a compression of the wage distribution: low wage rural workers have been replaced by low wage urban workers, and the wage hierarchy

---

5. See Piketty (2003).

remained more or less the same (in spite of the fact that real wages have been multiplied by 5 over the course of the century).

The fact that the drop in income inequality is solely due to the fall in top capital incomes, and that the fall took place mostly during wartime and the Great Depression, suggests an obvious explanation: for the most part, income inequality dropped because capital owners incurred severe shocks to their capital holdings during the 1914–1945 period (destructions, inflation, bankruptcies, etc.). This interpretation is confirmed by available wealth and estate data. Note that the idea that capital owners incurred large shocks during the 1914–1945 period and that this had a big impact on income distribution is certainly not new (Kuznets already mentioned this factor). What is new is that there is not much else going on.

The more challenging part that needs to be explained is the nonrecovery of top capital incomes during the post-1945 period (see Figure 1). Here the proposed explanation is that the 1914–1945 capital shocks had a permanent impact because the introduction of high income and estate tax progressivity (there was virtually no tax progressivity prior to 1914, and top rates increased enormously between 1914 and 1945) made it impossible for top capital holders to fully recover. Simple simulations suggest that the long-run impact of tax progressivity on wealth concentration is indeed large enough to explain the magnitude of the observed changes.[6]   **Fn 6**

The French case depicted on Figure 1 is interesting, because it appears to be fairly representative of what happened in other OECD countries.[7] In all countries   **Fn 7** for which we have data, the secular decline in income inequality took for the most part during the 1914–1945 period, and most of the decline seems to be due to the fall of top capital incomes. The 1914–1945 drop was larger in countries that were strongly hit by the war (France, Germany) than in the U.S., and there was no drop at all in countries not hit at all (such as Switzerland), which is consistent with the proposed explanation based on shocks. Moreover wealth concentration seems to have better recovered during the postwar period in countries with less tax progressivity (especially estate tax progressivity) such as Germany, which again seems broadly consistent with the tax explanation.

During the post-1970 period, one does observe a major divergence between rich countries. While top income shares have remained fairly stable in France and other Continental European countries over the past three decades, they have increased enormously in the U.S., where they are now back to their interwar levels (see Figure 2). The U.K. and other Anglo-Saxon countries tend be somewhere   **Figure 2** in between the European pattern and the U.S. pattern. Note that the rise of U.S. top income shares is not due to the revival of top capital incomes, but rather to the very large increases in top wages (especially top executive compensation).

---

6. See Piketty (2003).

7. See Atkinson (2005, this issue), Dell (2005, this issue), Saez (2005, this issue), and Atkinson and Piketty (2005).
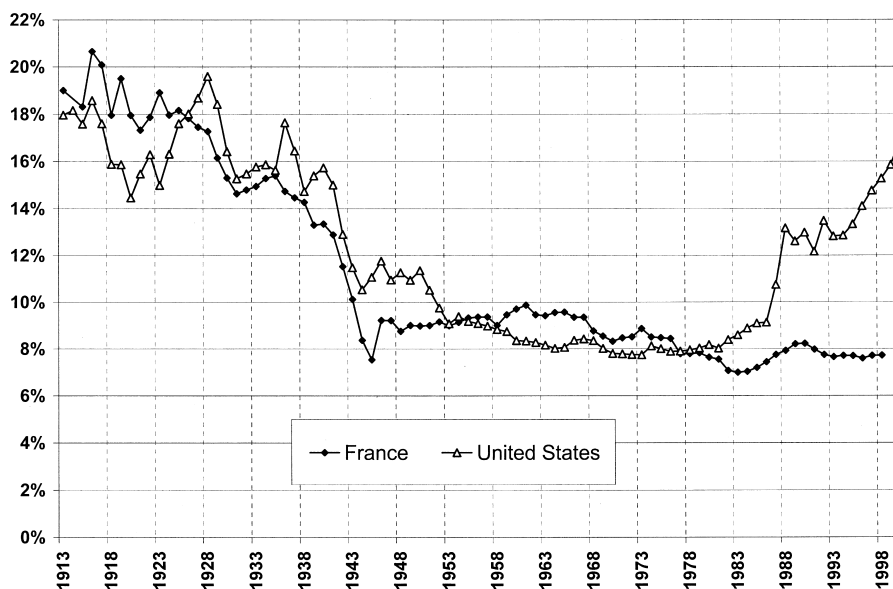
FIGURE 2. The top 1% income share in France and in the United States, 1913–2000. Source: Author's computations based on income tax returns (Piketty (2003) and Piketty and Saez (2003)).

As a consequence, top executives (the "working rich") have replaced top capital owners (the "rentiers") at the top of the U.S. income hierarchy over the course of the 20th century. This contrasts with the European pattern, where top capital incomes are still predominant at the top of the distribution (albeit at lower levels than at the beginning of the 20th century).[8] This provides yet another example as to why it is vital to be able to break down income distribution series by income source (without such a decomposition, it is virtually impossible to understand the forces at play). Note however the new U.S. pattern might not persist for very long: capital accumulation by the "working rich" is likely to lead the revival of top capital incomes in the following generation, especially in a context of large cuts in U.S. income and estate tax progressivity.

3. New Frontiers (I): Comeback to Cross-Country Structural Analysis

So far, most of the effort in our collective project has been devoted to constructing homogenous series and producing consistent analytical narratives as to why

---

8. See especially the striking contrast between the evolution of income composition patterns by fractile in the U.S. (Saez 2005, this issue, Figure 4) and Germany (Dell 2005, this issue, Figure 5).

income distribution evolved the way it did in the various countries. Although we believe one can learn a lot from carefully done case studies, the overall objective of the project is to provide a sufficiently rich database (with cross-country, temporal, and income source variations) so that one can conduct some rigorous cross-country testing of the various theoretical mechanisms at play. Although cross-country analysis will always suffer from severe identification problems, our hope is that richer data will allow to renew the analysis of the interplay between inequality and growth.

The first relationship that one might want to test in a systematic way is the impact of tax progressivity and other factors (such as fertility). Using standard stochastic models of capital accumulation, one can show that long run wealth concentration depends negatively on top estate tax rates and fertility:

$$b = F(t, n, \ldots)$$

With : $b = E(w|w > w_0)/w_0 = $ Pareto coefficient,
$t = $ top tax rate$(F_t < 0)$,
$n = $ fertility$(F_n < 0)$.

A high Pareto coefficient b means a fat upper tail of the distribution, i.e., high wealth concentration. Note that according to theoretical models tax progressivity and fertility should have an impact on the concentration of wealth and capital income, but not on the concentration of labor income. One can then calibrate these theoretical formulas to see whether observed differences in tax progressivity and fertility across countries can account for observed differences in wealth concentration. By going through such a calibration exercise, Dell (2005, this issue) concludes that relatively small differences in top estate tax rates can have a large impact on long-run wealth concentration. In particular, the difference in top estate tax rates between France and Germany appears to be large enough to account for the much higher concentration of wealth observed in Germany.

The other relationship that one might want to test using our database is the impact of inequality on growth. Several theories (e.g., the theory of credit constraints) predict that inequality might have a negative impact on growth. However the testing of these theories has been plagued by serious data problems. One could think of using our database to run standard cross-country regressions of the following sort:

$$\text{growth}_{it} = a + b \, \text{ineg}_{it} + \text{e}_{it}.$$

If one tries to run such regressions using our long run database (say, for France), the one would find a negative growth impact of inequality ($b < 0$). The reason is simply that the pre-1914 period (and to a large extent the interwar period) is associated to high inequality and relatively low growth, whereas the post-1945 period is associated to low inequality and high growth. Although we believe that

such regressions are more informative than standard cross-country-regressions on inequality and growth (our regressions rely on high-quality data and first-order changes in inequality), it is fairly obvious that this very crude methodology raises serious identification problems. There are lots of reasons why post-1945 growth was higher than pre-1914 growth (including a simple catching-up effect following the 1914–1945 shocks), and there is no way one can properly identify a causal impact of wealth concentration per se with such a crude regression. Using all countries in the database might allow to produce more convincing results. In the meantime, one can safely conclude that the enormous decline in wealth concentration that took place between 1914 and 1945 did not prevent high growth from happening.

4. New Frontiers (II): Beyond the Twentieth Century

The series constructed thus far focus for the most part upon the 20th century. As far as top income series are concerned, there is little hope to extend existing series to the 19th century. Modern income tax systems were introduced in most countries at the beginning of the 20th century, and there exists no systematic data source on incomes prior to this date. The prospects for extending existing series to the 19th century look better for the wealth distribution. In most Western countries, estate taxation or probate records have been in place for several centuries if one is ready to go back to individual-level administrative archives and collect large samples of wealth records, this material can be used to construct very long-run wealth distribution series. In France, a modern, universal estate tax was introduced in 1791, and individual estate tax returns have been stored and can be accessed in the local archives of each *departement*. When the estate tax became progressive in 1902, the tax administration started compiling and publishing tabulations by estate brackets. No such tabulation was compiled between 1791 and 1902, when the estate tax was purely proportional. In order to put our 20th century findings in perspective, we collected large samples of estate tax returns for all decedents with positive wealth in Paris every 10 years between 1807 and 1887, as well as a similar sample for 1902, in order to ensure the consistency of our 19th century series with the post-1902 tabulations-based series.[9]

As one can see from Figure 3, the basic finding is that wealth concentration in Paris and France kept rising right until World War 1. This is important, since this confirms that there was no preexisting, Kuznets-type trend in inequality priori to the 1914–1945 capital shocks. If anything, the upward trend in wealth concentration appears to accelerate at the end of the 19th century and at the beginning of the 20th century, which contradicts the Kuznets view of a stabilization or a reversal of the inequality trend after the initial wave of industrialization.

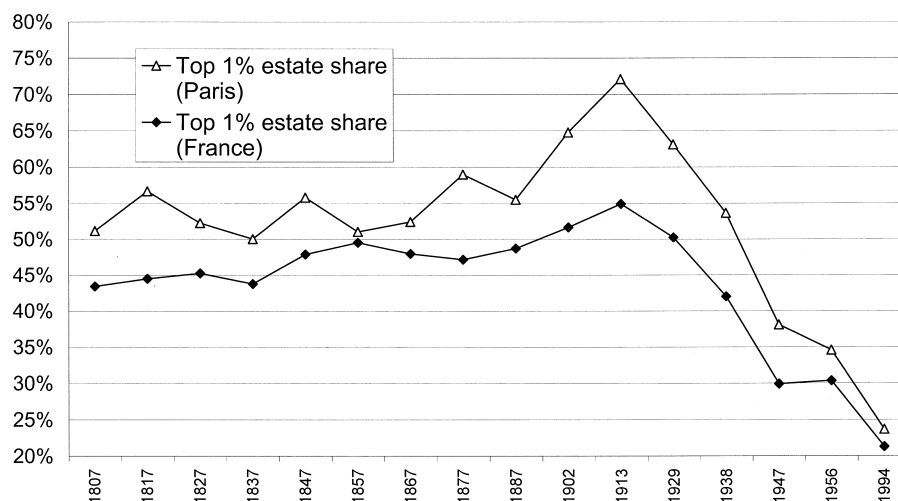9. See Piketty, Postel-Vinay, and Rosenthal (2004).

FIGURE 3. Wealth concentration in Paris and France, 1807–1994. Source: Authors' computations using estate tax returns (see Piketty, Postel-Vinay, and Rosenthal (2004)).

The fact that we have micro samples of estate tax returns (with detailed information on age, occupation, types of assets, etc.) also allows us to shed some new light regarding the impact of the inequality on growth. Per se, the existence of credit constraints does not necessarily imply that high wealth concentration is bad for growth. If most of the wealth is owned by active entrepreneurs who keep reinvesting their assets in profitable projects, high wealth concentration is not necessarily bad. However if most of the wealth is owned by retired rentiers investing their wealth in low-yield assets, then high wealth concentration can entail substantial efficiency costs. Here the striking finding is that wealth was getting older and older in France during the 19th century and until World War I (see Table 1). There is also evidence that top wealth holders were investing a rising fraction of their wealth in low-yield assets such as public bonds. Although this is not sufficient to prove that inequality had a negative growth impact, this shows that the very high levels of wealth concentration that prevailed in France at the eve of World War 1 were associated to retired rentiers rather than to active entrepreneurs (with potential damaging growth effects).

Table 1

5. New Frontiers (III): Beyond the OECD

Most of our series are devoted to OECD countries. The reason is simply that many LDCs introduced a modern income tax only recently, so it is generally impossible to construct long-run income distribution series for these countries.

TABLE 1. The age profile of wealth at death in Paris, 1817-1994 (average estate left by 50–59 yr-old = 100).

	20-29-yr-old	30-39-yr-old	40-49-yr-old	50-59-yr-old	60-69-yr-old	70-79-yr-old	80-89-yr-old	90-99-yr-old
1817	48	49	49	100	80	70	70	
1827	49	46	73	100	94	99	63	
1837	67	79	107	100	112	123	102	
1847	78	73	102	100	117	154	135	
1857	78	77	101	100	104	102	111	
1867	65	54	82	100	132	141	142	
1877	56	62	63	100	205	262	340	
1887	36	27	66	100	130	214	288	
1902	30	40	80	100	169	239	251	
1947	31	51	73	100	113	105	105	109
1994		11	45	100	87	93	95	68

Source: Author's computations using estate tax returns (see Piketty, Postel-Vinay, and Rosenthal 2004).

There are however some exceptions. For instance, a progressive income tax was introduced in 1922 in India, which allowed us to compute 1922–2000 top income share series for India.[10] More generally, we believe that even in LDCs where the income tax was introduced recently, income tax return data should be used as a useful supplement to standard income surveys. In particular, one problem with standard surveys is that they severely underestimate top incomes (this is true everywhere, but especially so in LDCs). The Indian "growth paradox" of the 1990s provides us with an example to illustrate this point. As measured by national accounts, Indian growth was very high during the 1990s. However, there was much lower growth according to household expenditure surveys. One possible explanation is the rise of top incomes (not properly recorded in surveys). By using income tax returns data, we were indeed able to show that the rise in the top percentile income share alone can account for between 20% and 40% of this "growth paradox". Cuts in top tax rates in India during the 1990s were relatively modest, so the substantial rise in reported top incomes probably reflects a real economic phenomenon (rather than pure fiscal manipulation).[11] We hope that tax return data will prove to be useful to study income inequality dynamics in other low- and middle-income countries in the future.

References

Atkinson, Anthony (2005). "Comparing the Distribution of Top Incomes across Countries." *Papers and Proceedings 2004, Journal of the European Economic Association* 3(2–3) this issue.

---

10. See Banerjee and Piketty (2004).

11. Moreover, note that top income shares also rose in China during the 1990s (about twice as much as in India), in spite of the fact that there was no top tax cut at all (see Piketty and Qian 2004).

Atkinson, Anthony and Andrea Brandolini (2001). "Promise and Pitfalls in the Use of Secondary Data-Sets: Income Inequality in OECD Countries as a Case Study." *Journal of Economic Literature*, 39, 771–799.

Atkinson, Anthony and Thomas Piketty) (forthcoming). *Top Incomes over the Twentieth Century*, Oxford University Press.

Banerjee, Abhijit and Thomas Piketty (2004). "Top Indian Incomes, 1922–2000." CEPR Discussion Paper 4632.

Dell, Fabien (2005). "Top Incomes in Germany and Switzerland over the Twentieth Century." *Papers and Proceedings 2004, Journal of the European Economic Association* 3(2–3) this issue.

Kuznets, Simon (1953). *Shares of Upper Income Groups in Income and Savings*. National Bureau of Economic Research.

Kuznets, Simon (1955). "Economic Growth and Economic Inequality." *American Economic Review*, 45(1), 1–28.

Piketty, Thomas (2001). *Les hauts revenues en France au XXe siècle—Inégalités et redistributions*, 1901–1998, Grasset, Paris.

Piketty, Thomas (2003). "Income inequality in France, 1901–1998." *Journal of Political Economy*, 111, 1004–1042.

Piketty, Thomas Gilles Postel-Vinay and Jean-Laurent Rosenthal (2004). "Wealth Concentration in a Developing Economy: Paris and France, 1807–1994." CEPR Discussion Paper 4631.

Piketty, Thomas and Nancy Qian (2004). "Income Inequality and Progressive Income Taxation in China and India, 1986–2010." Working Paper, Paris- Jourdan and MIT.

Piketty, Thomas and Emmanuel Saez (2003). "Income Inequality in the United States, 1913-1998." *Quarterly Journal of Economics* 118, 1–39.

Saez, Emmanuel (2005). "Top Incomes in the United States and Canada over the Twentieth Century." *Papers and Proceedings 2004, Journal of the European Economic Association* 3(2–3), this issue.