1

Top Incomes Over the Twentieth Century: A Summary of Main Findings[1]

T. Piketty

1.1 INTRODUCTION

This introductory essay presents some of the key findings and perspectives emerging from the detailed country chapters published in this volume. All chapters are part of a collective research project on the long-run dynamics of income and wealth distribution. The general 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 twentieth century for over 20 (mostly Western) countries. The present volume focuses upon the contrast between continental European countries and English-speaking countries and includes ten case studies: France, UK, US, Canada, Australia, New Zealand, Germany, the Netherlands, Switzerland, and Ireland. A forthcoming volume will complete the study by covering Scandinavian and Northern Europe (including Sweden, Finland, and Norway), Southern Europe (including Italy, Spain, Portugal), as well as a number of Latin American (including Argentina, Brazil) and Asiatic countries (including India, China, and Indonesia).

The primary motivation for this project was a general dissatisfaction with existing income distribution databases. The international databases on inequality that existed were not high quality (they display little homogeneity over time or across countries),[2] they are not long-run (typically they 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 labour income and a capital income component. This latter feature of existing data sets is unfortunate, because the economic mechanisms at work can be very

---

[1] The references to this chapter are given at the end of Chapter 2.

[2] See, e.g., the Atkinson-Brandolini (2001) criticism of the World Bank (Deininger-Squire) secondary database. The 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.

different for the distribution of labour income (demand and supply of skills, labour 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.[3]

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), and fully homogenous, cross-country data sets do not exist. However, our database has the following advantages:

- we use the same raw data sources for all countries and apply the same methodology to derive the final series;
- the series are typically annual and cover a long-run of years;
- the data are mostly broken down by income source.

This means that they offer a unique opportunity to understand better the dynamics of income and wealth distribution and the two-way interaction between inequality and growth.

We should stress that the main objective of the chapters collected in this volume is to describe how the series were constructed, and to offer first cut analysis of the long-run dynamics of inequality in each individual country. Such analytical narratives and detailed case studies are useful, but in our view they should be seen as complements (rather than substitutes) to a more systematic statistical exploitation of the complete database, which we do not offer in this volume. We very much hope that future researchers will use our database to explore causal mechanisms in a more systematic way, and in particular that our data will contribute to renew the literature on cross-country inequality/growth regressions.[4]

The rest of this introductory essay is organized as follows. In section 1.2, we briefly present the basic data and methodology used to construct the database. Section 1.3 presents some of the main descriptive findings and conclusions, with particular emphasis to the Kuznets' curve debate. Section 1.4 attempts to illustrate how our database could potentially be used to renew the cross-country structural 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 the database using additional published historical tax tabulations and collecting historical individual tax data (Section 1.5).

---

[3] This was first stressed by Kuznets (1955).

[4] One of the key reasons why the literature on cross-country inequality/growth regressions failed to deliver robust conclusions (see, e.g., Banerjee and Duflo (2003) for a critical appraisal) is the poor quality of existing databases.

1.2. CONSTRUCTING A NEW DATABASE: PRIMARY DATA AND METHODOLOGY

Household income surveys are a relatively recent venture: they virtually did not exist on a national basis prior to 1950, and in most countries they are not available in a 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 twentieth century (1913 in the US, 1914 in France, etc.), and in all countries with an income tax system the tax administration started compiling and publishing tabulations based on the exhaustive set of income tax returns.[5] These tabulations generally report for a large number of income brackets the corresponding number 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.

In order to give a sense of what our primary data sources look like, we reproduce in Table 1.1 the raw top income tabulations for France in 1919, as they were originally published by the Finance Ministry. One can see for instance on this table that 181 French taxpayers reported tax income above one million francs in 1919 (a pretty large income at that time). We also reproduce on Table 1.2 the raw income composition tabulations for France in 1920. One can see that out of the 722 million French francs reported by French taxpayers with individual income above 1 million francs in 1920, 322 million francs took the form of 'revenus des valeurs et capitaux *mobiliers*' (interest and dividend income), 356 million francs took the form of '*bénéfices industriels et commerciaux*' (business income), and only 16 million francs took the form of '*traitements publics et privés, salaires, etc.*' (wage income).

One can then use standard Pareto extrapolation techniques to compute top fractiles thresholds and average incomes using such data. This methodology is described in a detailed manner in Chapter 2. Here it is sufficient to recall that the Pareto law for top incomes is given by the following distribution function:

$$1 - F(y) = (k/y)^a \ (k > 0, \ a > 1) \tag{1.1}$$

The corresponding density function is given by $f(y) = ak^a/y^{(1+a)}$. The key property of Pareto distributions is that the ratio between the average income y^* (y) of individuals (or households or tax units) with income above y and y does not depend on the income threshold y:

$$y^*(y) = \left[\int_{z>y} zf(z)dz \right] / \left[\int_{z>y} f(z)dz \right]$$

$$= \left[\int_{z>y} dz/z^a \right] / \left[\int_{z>y} dz/z^{(1+a)} \right] = ay/(a-1) \tag{1.2}$$

i.e. $y^*(y)/y = b$, *with* $b = a/(a-1)$

---

[5] Full details about the administrative publications where the raw tabulations were originally published are given in the country chapters.

Table 1.1 Raw top income tabulations, France 1919 (IMPÔT GÉNÉRAL SUR LE REVENU)

CATÉGORIES DE REVENUS.	NOMBRE de CONTRIBUABLES inscrits dans les rôles.	MONTANT des REVENUS imposés.	MONTANT DES DÉDUCTIONS			MONTANT BRUT de l'impôt.	MONTANT des PÉNALITÉS et droits au sus.
			pour SITUATION de famille.	pour CHARGES DE FAMILLE			
				1,500 fr.	5,000 fr.		
1	2	3	4	5	6	7	8
		fr.	fr.	fr.	fr.	fr.	fr.
6,100 à 10,000 francs	130,787	1,170,324,800	123,915,000	7,110,000	34,406,000	3,805,400	170,500
10,100 à 20,000	193,679	2,851,910,400	417,507,000	25,410,000	194,082,000	21,056,000	759,100
20,100 à 30,000	58,894	1,477,045,800	137,517,000	8,983,500	97,740,000	18,687,300	755,300
30,100 à 50,000	39,974	1,529,512,700	93,711,000	6,235,500	79,134,000	40,061,400	1,025,200
50,100 à 100,000	23,882	1,592,572,500	62,733,000	3,354,000	46,894,000	94,486,600	1,907,700
100,100 à 200,000	9,487	1,517,031,000	21,768,000	1,513,500	50,530,000	142,413,800	2,820,500
200,100 à 300,000	2,289	556,396,900	6,651,000	315,000	5,456,000	99,524,900	965,900
300,100 à 500,000	1,388	527,734,800	3,204,000	138,000	3,080,000	126,024,700	1,228,500
500,100 à 1 million	576	387,082,900	1,380,000	46,500	1,318,000	130,956,900	1,680,800
Au-dessus de 1 million	181	451,968,100	420,000	13,500	336,000	206,785,300	883,400
TOTAUX	467,137	11,867,588,900	868,911,000	53,119,500	492,776,000	883,801,200	12,177,000

Table 1.1 (Contd.)

	MAJORATION DUE PAR LES CONTRIBUABLES CÉLIBATAIRES. (25 p. 100.)			MAJORATION DUE PAR LES MÉNAGES SANS INFANTS. (10 p. 100.)				
Nombre de contribuables supportant la majoration.	Montant des revenus des intéressés	Produit de la majoration.	Nombre de contribuables supportant la majoration.	Montant des revenus des intéressés.	Produits de la majoration.	MONTANT des DÉDUCTIONS pour charges de famille.	PRODUIT NET TOTAL de l'impôt.
9	10	11	12	13	14	15	16
	fr.	fr.		fr.	fr.	fr.	fr.
45,190	340,334,700	430,700	11,900	111,048,300	70,000	105,600	4,570,800
21,602	301,518,900	875,600	29,401	413,678,300	354,800	727,800	22,518,600
5,162	130,728,500	1,026,200	6,712	168,608,500	470,500	096,200	20,225,100
3,398	132,038,600	1,073,900	4,225	168,148,800	431,000	1,801,200	40,777,300
2,049	143,370,600	2,067,800	2,407	168,390,600	1,034,800	5,028,800	95,868,100
746	99,947,600	3,173,800	904	125,934,300	1,557,000	10,161,500	139,805,600
167	39,950,600	1,886,300	196	46,776,500	893,000	4,704,100	98,566,000
114	33,245,200	2,153,300	123	45,315,200	1,137,000	3,080,000	127,403,800
35	24,508,500	2,087,500	45	29,941,900	1,086,100	1,518,000	134,493,500
23	49,247,600	5,993,800	17	33,763,300	1,506,700	336,000	214,833,200
TOTAUX 78,492	1,294,870,800	20,770,900	55,930	1,511,005,700	8,590,900	28,620,200	896,719,800

Note: TABLEAU *présentant, à la date du 30 avril 1922, la décomposition, par catégories de revenus, des résultats des rôles établis au titre de l'année 1920 (revenus de 1919)*

Source: Originally published in Bulletin de statistique et de législation comparée, March 1923: vol. 93.

Table 1.2 Raw income composition tabulations, France 1919 (IMPÔT GÉNÉRAL SUR LE REVENU.)

DÉCOMPOSITION DES REVENUS GLOBAUX SUIVANT LES DIVERSES SOURCES D'OÙ ILS PROVIENNENT (a).

CATÉGORIES DE REVENUS	MONTANT TOTAL des revenus globaux (A)	REVENUS des propriétés bâties		REVENUS des propriétés non bâties		REVENUS des valeurs et capitaux mobiliers.		BÉNÉFICES de l'exploitation agricols.	
	Montant	Montant.	Proportion.	Montant.	Proportion.	Montant.	Proportion.	Montant.	Proportion.
1	2	3	4	5	6	7	8	9	10
	millions.	millions.	%	millions.	%	millions.	%	millions.	%
6,100 à 10,000 fr....	1,100	65	5.9	31	3.1	148	13.5	18	1.6
10,100 à 20,000	3,832	205	5.3	100	2.6	497	13.0	82	2.1
20,100 à 30,000	2,044	127	0.3	63	3.1	301	17.7	47	2.3
30,100 à 50,000	2,132	142	6.7	62	2.9	402	21.7	40	1.9
50,100 à 100,000	2,281	143	0.3	59	2.6	586	25.0	36	1.6
100,100 à 200,000	1,803	97	5.4	30	1.7	514	28.5	18	1.0
200,100 à 300,000	751	34	4.5	10	1.3	233	31.0	5	0.7
300,100 à 500,000	699	20	3.7	8	1.1	227	32.5	6	0.9
500,100 à 1 million...	530	17	3.2	4	0.7	186	35.1	4	0.8
Au-dessus de 1 million...	722	12	1.7	5	0.7	322	44.6	3	0.4
TOTAUX ET MOYENNES	16,897	868	5.5	375	2.4	3,536	22.2	259	1.6

Table 1.2 (*Contd.*)

	BÉNÉFICES Industriels et commerciaux.		BÉNÉFICES de l'exploitation minière.		TRAITEMENTS publics et privés, salaires, etc.		PENSIONS de rentes viagères		BÉNÉFICES des professions non commerciales.		REVENUS des charges et offices.	
	Montant.	Proportion.	Montant.	Proportion.	Montant.	Proportion.	Montant.	Proportion.	Montant.	Proportion.	Montant.	Proportion.
	11	12	13	14	15	16	17	18	19	20	21	22
	millions.	%	millions.	%	millions.	%	millions.	%	millions.	%	millions.	%
	107	9.7	1	0.1	665	60.4	35	3.2	24	2.2	3	0.3
	820	21.6	3	0.1	1,873	48.9	72	1.9	148	3.8	26	0.7
	651	31.8	3	0.1	634	31.0	29	1.4	102	5.0	27	1.3
	705	35.0	2	0.1	502	23.5	22	1.0	101	4.7	34	1.6
	800	39.0	2	0.1	422	13.5	17	0.7	92	4.0	37	1.6
	800	44.4	3	0.2	257	14.2	8	0.4	50	3.1	20	1.1
	353	47.0	1	0.1	87	11.0	3	0.4	18	2.4	7	1.0
	314	49.2	1	0.1	68	9.7	3	0.3	14	2.0	3	0.5
	200	50.2	1	0.2	43	8.1	2	0.2	8	1.5	"	"
	356	49.3	3	0.4	16	2.2	"	"	5	0.7	"	"
TOTAUX ET MOYENNES	5,358	33.7	20	9.1	4,567	28.7	189	1.2	568	3.6	167	1.0

Notes: IMPÔT ÉTABLI AU TITRE DE L'ANNÉE 1921,—BÉNÉFICES ET REVENUS RÉALISÉS AU COURS DE L'ANNÉE 1920. *Tableau présentant, pour les contribuables inscrits dans les rôles émis du 1st janvier 1921 au 30 avril 1922, la décomposition du revenu global (Revenus déclarés seulement) les différentes sources de revenus.*

(a) Avant toute déduction au titre des charges grovant le revenu global. (Contributions directes assimilées, pertes résultant d'un déficit d'exploitation, intérêts de dettes, etc.).

(b) Aucune concordance ne peut exister entre le montant de revenus indiqués au présent tableau et le montant des revenus quel servi de base aux impôts cédulaires pour l'anée 1921. Tous les contribuables assujettis aux impôts cédulaires ne sont pas, en effet, *possibles* de l'impôt général et, inversement, certains revenus entrant dans la composition du revenu global soumi à l'impôt général ne sont pas frappés par l'impôt cédulaire parce que leur montant ne dépense pas la somme affranchie de l'impôt dans la cédule correspondante.

Source: Originally published in *Bulletin de statistique et de législation comparée*, March 1923: vol. 93.

That is, if $b=2$, the average income of individuals with income above €100,000 is €200,000, and the average income of individuals with income above €1 million is €2 million. Although this law is only an asymptotic approximation (in practice, estimated b coefficients vary slightly with y), it works remarkably well for top incomes, as was first noted by Vilfredo Pareto (1896, 1896–97) in the 1890s using tax tabulations from Swiss cantons. In this volume, we do not address the interesting issue as to why this law holds, and we solely use is as an interpolation technique allowing us to compute top fractile thresholds and average incomes from grouped income data. It is important to note that although the b coefficient is (almost) invariant with y for a given country and a given year, it does vary substantially over time and across countries.[6] A higher b coefficient means a fatter upper tail of the income distribution, which generally implies higher inequality (for a constant mean). For instance, the b coefficient declined from about 2.3–2.4 to about 1.7–1.8 in France during the twentieth century, as top income shares dropped. The b coefficient went through a similar decline in all countries where inequality dropped, and it started rising again in countries where inequality rose since the 1970s, e.g. in the United States (where the b coefficient is now back to about 2.3–2.4).

Pareto extrapolation techniques are fairly powerful, but they do not allow extrapolation on income ranges for which we have no data. In that respect, one major limitation of tax data is that the income of individuals not subject to the tax is excluded from the data. Prior to the Second World War, 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 fractile income shares series from top fractile income data, one needs a total income denominator, which can be computed using aggregate income sources (national accounts and their ancestors). Constructing homogenous numerator and denominator series requires special care and raises a number of issues, many of which are addressed in Chapter 2.

1.3 BASIC DESCRIPTIVE FINDINGS: THE KUZNETS' CURVE, 50 YEARS LATER

The first economist to use these data sources and methodology in a systematic way was Kuznets (1953).[7] He exploited US income tax tabulations covering the

---

[6] Most authors refer to $a = b/(b-1)$ (rather than b) as the 'Pareto coefficient'. Note, however, that the b coefficient has a more intuitive economic meaning. One could for instance refer to $b-1$ as the 'income advantage of the rich' (IAR) coefficient. During the twentieth century the IAR coefficient declined from 130–140% to 70–80% in France, i.e. the income advantage of the rich nearly halved.

[7] Earlier authors (e.g. Bowley 1914 and Stamp 1916) used income tax data in a sophisticated way (see Chapter 4), but Kuznets was apparently the first scholar to use control totals to construct top income shares series.

1913–48 period and computed corresponding top decile and top percentile income shares series. These were the first long-run income distribution series ever produced (income distribution had been at the centre of speculative economic thought at least since the time of Ricardo and Marx, but few data were available). Unsurprisingly, these series had a major impact on economic thinking, especially after Kuznets (1955) proposed his famous 'Kuznets curve' theory in order to account for the 1913–48 decline in income inequality that he witnessed for the United States. According to this theory (which Kuznets himself viewed as highly speculative),[8] 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.

In a sense, all what we are doing in this project 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 instead of 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 than Kuznets as to why income inequality dropped in Western countries during the first half of the twentieth century. First, as one can see on Figure 1.1, where we plot the basic series for the French case, the decline in top income shares witnessed by Kuznets for the US also took place in France, but it came to an end right after the Second World War. The secular decline in income inequality took place during a very particular and politically chaotic period, namely during the 1914–45 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 particular.

Next, and most importantly, one can see from Figure 1.1 that the 1914–45 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 twentieth 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); see Piketty (2003) and Chapter 3. Labour

---

[8] 'This is perhaps 5% empirical information and 95% speculation, some of it possibly tainted by wishful thinking' (Kuznets 1955: 26).
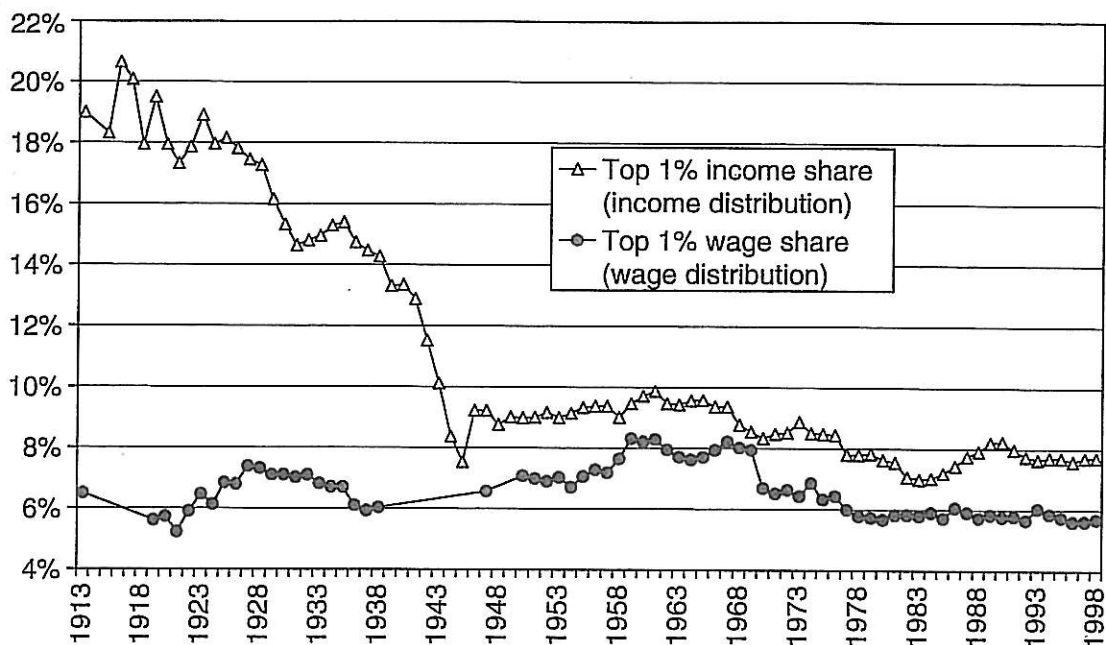
T. Piketty

Figure 1.1 The fall of top capital incomes in France, 1913–98

Source: Piketty 2001, 2003, Chapter 3 this volume: Table 3A.1; authors' computations using income tax returns.

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 twentieth century, and rural workers have virtually disappeared by the end of the twentieth 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 remained more or less the same (in spite of the fact that real wages have been multiplied by five 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 1914–45 period (destruction, 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 1914–45 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 non-recovery of top capital incomes during the post-1945 period (see Figure 1.1). Here the proposed explanation is that the 1914–45 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 (Piketty (2003) and Chapter 3) suggest that the

long-run impact of tax progressivity on wealth concentration is indeed large enough to explain the magnitude of the observed changes.[9]

The French case depicted on Figure 1.1 is interesting, because it appears to be fairly representative of what happened in other OECD countries.[10] In all countries for which we have data, the secular decline in income inequality took place for the most part during the 1914–45 period, and most of the decline seems to be due to the fall of top capital incomes. The 1914–45 drop was larger in countries that were strongly hit by the war (e.g. France and Germany) than in the US, 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 capital shocks. Moreover wealth concentration seems to have better recovered during the post-war period in countries with less tax progressivity (especially estate tax progressivity) such as Germany, which again seems broadly consistent with the tax explanation.

There are however important differences between rich countries. First, income inequality did keep declining during the 1950s–1960s in a number of countries (such as the UK), albeit at a lower pace than during the 1914–45 period.[11] Next, 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 US, where they are now back to their interwar levels (see Figure 1.2). The UK and other Anglo-Saxon countries tend be somewhere in between the European pattern and the US pattern. Note that the rise of US 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). As a consequence, top executives (the 'working rich') have replaced top capital owners (the 'rentiers') at the top of the US income hierarchy over the course of the twentieth 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 twentieth century).[12] 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 US pattern might not persist for very long: capital accumulation by the 'working rich' is likely to lead the revival of top capital incomes at the following generation, especially in a context of large cuts in US income and estate tax progressivity.

Although most countries covered in this volume do follow this general pattern (abrupt decline of top capital incomes during the 1914–45, sudden rise of top wages in Anglo-Saxon countries since the 1970s), a careful reading of the country chapters collected in this volume will reveal many interesting particularities.

[9] See Piketty (2003) and Chapter 3 in this volume.

[10] See the country chapters collected in this volume.

[11] This might be partly due to the steeply progressive tax structure applied in those countries (especially in the UK), but there are other explanations as well.

[12] See especially the striking contrast between the evolution of income composition patterns by fractile in the US (Saez 2005: fig. 4) and Germany (Dell 2005: fig. 5).
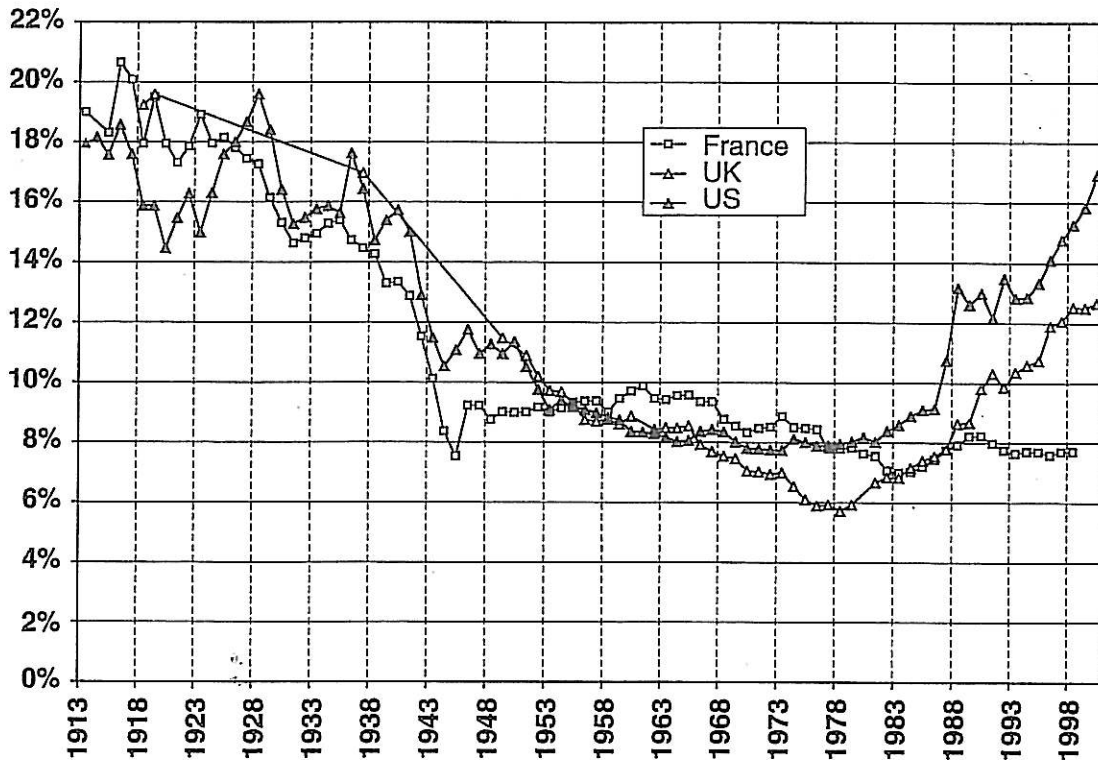
Figure 1.2 The top 1% income share in France, the UK, and the US, 1913–2000

Source: France: Piketty Chapter 3 (this volume): Table 3A.1; UK: Atkinson Chapter 4 (this volume): Table 4.1; US: Piketty and Saez Chapter 5 (this volume): Table 5A.1; authors' computations using income tax returns.

We already mentioned the very special case of Switzerland, where top shares have been basically flat in the long run. Countries like Ireland, Australia, and New Zealand, which were less strongly affected by the wars than other countries, also witnessed a limited inequality decline during the 1914–45 period (albeit less limited than in Switzerland, for reasons that probably have to do with differences in trade structures with countries at war). Top income shares in Canada have increased dramatically since the 1970s, thereby confirming the existence of a distinct Anglo-Saxon pattern, as opposed to continental Europe (e.g. France, Germany, and the Netherlands), where top shares hardly changed during the past 30 years. The case of Germany reveals another interesting pattern: although top German capital incomes were strongly hit by the Second World War, they seem to have recovered fairly quickly and to be structurally higher than in other Western countries, for reasons that might be related to the limited tax progressivity of the German fiscal system (more on this below).

1.4. NEW FRONTIERS (I): RETURN 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

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 a renewal of 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 capital income or wealth concentration depends negatively on top income and estate tax rates and fertility:

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

Where $b = E(w|w > w_0)/w_0 = IAR$ (Pareto) coefficient, $\qquad$ (1.3)
t = top tax rate ($G_t < 0$),
n = fertility ($G_n < 0$)

A high 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 labour income. One can then calibrate these theoretical formulae 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) 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 data base 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 data base to run standard cross-country regressions explaining the growth rate of country I at time t as a function of the inequality in country I at time t. If one tries to run such regressions using our long-run data base (say for France), then one would find a statistically significant, negative growth impact of inequality. 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 1914–45 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 data base might allow production of more convincing results.[13] 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.

1.4 NEW FRONTIERS (II): EXTENDING THE INEQUALITY DATABASE

Although the international long-run inequality data base presented in this collective volume covers a large number of years and countries, it is far from being complete. First, historical income tax tabulations do exist for many more countries than the ten countries covered in the present volume, and the companion volume will include additional countries in Scandinavia and Northern Europe, Southern Europe, Latin America, and Asia. More countries are yet to be explored, both in the OECD and in the developing world. Note that our long-run data base is bound to be devoted for the most part to OECD countries. One reason is simply that a number of LDCs introduced a modern income tax only recently, so it is often impossible to construct long-run income distribution series for these countries. There are, however, some exceptions. For instance, a progressive income tax was introduced in 1922 in India, which allows the computation of the 1922–2000 top income share series for India (Banerjee and Piketty 2005). In addition to the countries covered in the companion volume, there probably exist a number of other non-OECD countries (especially ex-colonies) where tax data spanning reasonably long time periods are available. Note that even in LDCs where the income tax was introduced only recently, income tax returns data should probably be used more often as a useful supplement to standard income surveys.[14]

Next, the series constructed for the ten countries covered in the present volume are incomplete, in the sense that an exhaustive use of all published tax tabulations in these countries would allow the construction of a number of additional series. For all countries, we offer annual homogenous series on top income shares

---

[13] For a first attempt to use the data base to conduct panel cross-country regressions, see Atkinson and Leigh (2004) and Leigh (2006).

[14] For instance, it is only in 1980 that a modern progressive income tax was introduced in China (following the 1979 reforms), so that it is impossible to construct long-run Chinese inequality series. However, Chinese tax data available for the 1980s–90s offers a useful supplement to standard surveys, e.g. in order to compare inequality dynamics in China and India during the recent period (see Piketty and Qian 2004). In particular, one problem with standard surveys is that they severely under-estimate top incomes (this is true everywhere, but especially so in LDCs), and tax data allows us to address puzzling facts such as the Indian 'growth paradox' of the 1990s (see Banerjee and Piketty 2005).

covering most of the twentieth century. However, available tax tabulations also allow us to calculate effective income tax rates series for each top income fractile. This is a fairly tedious work (this requires collecting exhaustive information on tax law and taking into account all variations in family structure, children allowances, etc.), and such series have been constructed for only a handful of countries.[15] Available income composition data was used for most countries covered in this volume, albeit not always on an annual basis.[16] In countries with a progressive estate tax, there also exists a whole set of historical estate tax tabulations, which could be used to compute top estate shares series (wealth distribution among decedents), as well as top wealth shares series (wealth distribution among the living) using the estate multiplier.[17] In the context of this volume, we chose to concentrate on income tax tabulations and top income shares series, and we did not attempt to use estate tax tabulations in a systematic way.[18] Extending the data base in this direction raises technical difficulties but would be a useful step in order to enrich cross-country structural regressions.

Finally, and most importantly, one of the most exciting avenues for extending historical inequality data sets in the future probably consists of collecting micro-level tax data from individual tax returns available in national archives. As this volume attempts to illustrate, published tax tabulations are a useful date source and allow us to gain a better understanding of the long-run determinants and consequences of income inequality. However it is obvious that one could do a lot more if micro-level data sets were available. In most OECD countries, micro-level tax returns data sets are available only for the post-1970 or post-1980 period, and they usually cover a limited number of years and use a fairly low sampling rate.[19] The only way to construct micro-data sets for earlier periods and with adequate sampling rate is to go back to individual tax returns stored in national archives

[15] Note that available data on family structure, number of children, etc., for each income bracket could also be used to study marriage and fertility behaviour for each top income fractile and to analyse the behavioural impact of changing financial incentives.

[16] In some countries (e.g. France and the US), separate tabulations by wage brackets were also published and have been used to compute top wage shares series (and not only top income shares and top income composition series).

[17] In countries with a comprehensive tax on the wealth of the living (this is less common than a comprehensive estate tax), the corresponding data can also be used to compute top wealth shares series.

[18] Estate tax tabulations were used in a systematic way by Atkinson and Harrison (1978) for the UK (earlier authors did use estate tax data to produce top wealth shares estimates, albeit for shorter periods; official top wealth shares are now published every year by the UK Inland Revenue) and by Lampman (1962) (the resulting top wealth shares series have recently been extended until the present day for the US by Kopczuk and Saez (2004)). Similar series are also available for France (see Piketty 2001, 2003; Chapter 3; and Piketty et al. 2004). The chapter on Switzerland (Chapter 11) also uses wealth data, although not in a systematic way.

[19] One exception is the US, where the Internal Revenue Service (IRS) released annual micro-level data sets for income tax returns starting in 1960 and with large over-sampling at the top (see Piketty and Saez 2003 and Chapter 5 in this volume). In most countries, micro-level data sets with large over-sampling at the top (or sometime exhaustive date sets) have been used by tax authorities since the 1970s but are difficult to access for researchers.

(older returns were destroyed in some countries, but properly stored in others) and scan hundreds of thousands of them. Depending on technological evolution and financial resources made available for such projects, scholars working on historical changes in income distribution might throw away tax tabulations and start working on long run micro-level tax returns data set in ten years, 50 years, or more.

In order to illustrate what micro-level data sets could bring to the analysis of historical changes in inequality, we take the example of a recent study on wealth concentration in Paris and France over the 1807–1994 period. 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 *département*. 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 twentieth century top wealth shares series in perspective, Piketty et al. (2004) collected large samples of estate tax returns for all decedents with positive wealth in Paris every ten years between 1807 and 1887, as well as a similar sample for 1902, in order to ensure the consistency of the nineteenth-century series with the post-1902 tabulations based series. As one can see from Figure 1.3, the basic finding is that wealth concentration in Paris and France kept rising right until the First World War. This is important, since this confirms that there was no pre-existing, Kuznets-type trend in inequality prior to the 1914–45 capital shocks. If anything, the upward trend in wealth concentration appears to accelerate at the end of the nineteenth century and at the beginning of the twentieth century, which again
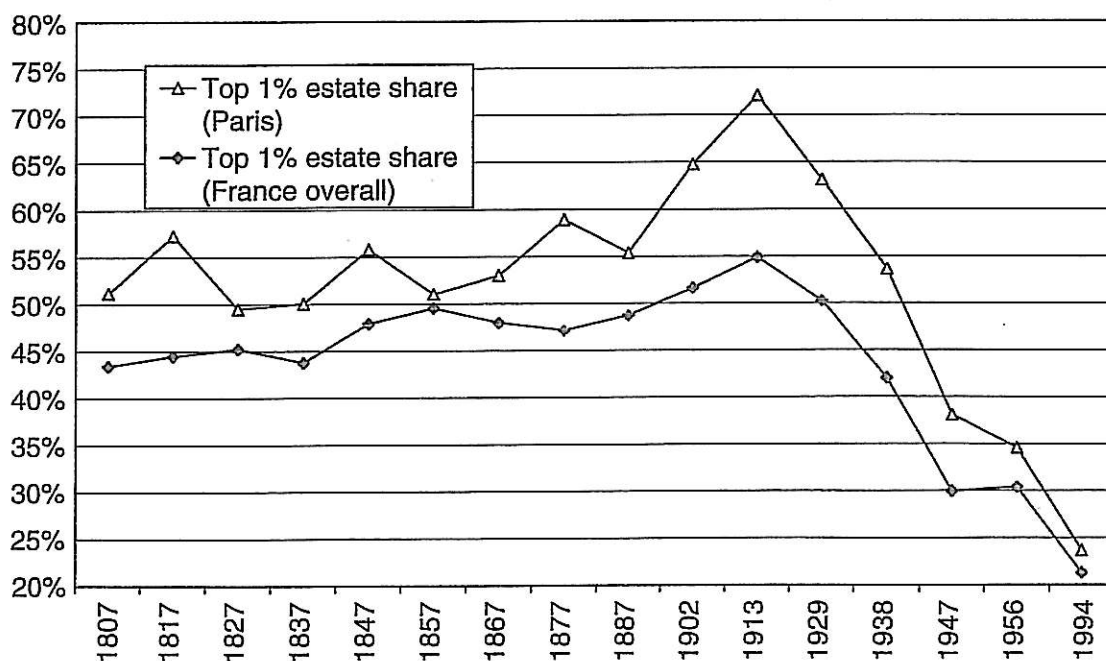

Figure 1.3 Wealth concentration in Paris and France overall, 1807–1994

Source: Piketty et al. 2004; authors' computations using estate tax returns.

Table 1.3 The age profile of wealth at death in Paris, 1817–1994

	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	26	22	28	100	54	59	59	—
1827	44	50	53	100	88	87	60	—
1837	133	90	107	100	116	123	110	—
1847	87	73	102	100	117	204	132	—
1857	84	77	101	100	104	109	145	—
1867	67	58	136	100	141	125	154	—
1877	66	73	63	100	197	260	430	—
1887	45	33	63	100	152	233	295	—
1902	29	40	80	100	253	272	401	—
1947	31	51	73	100	113	105	105	109
1994	—	11	45	100	87	93	95	68

Note: Average estate left by 50–59-yr-old = 100.

Source: Piketty et al. 2004; authors' computations using estate tax returns.

contradicts the Kuznets view of a stabilization or a reversal of the inequality trend after the initial wave of industrialization.

Most importantly, the fact that we now 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 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 re-investing 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 nineteenth century and until the First World War (see Table 1.3). 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 the First World War were associated with retired rentiers rather than with active entrepreneurs (with potential damaging growth effects). The data set also makes it possible to study the evolution of the share of aristocratic fortunes, to test hypothesis about the changing share of women in top wealth fractiles, etc.[20] With sufficient resources one could also construct panel data sets and follow the same individuals or dynasties over time. If and when such data sets become available for a large number of countries, both for income and estate tax returns, the scientific study of income distribution will take a new turn. But in the meantime, we very much hope that this volume will convince the reader that a systematic use of published tax tabulations allows us to make progress in this direction.

[20] See Piketty et al. (2004) for a detailed analysis.