CHAPTER 15

Wealth and Inheritance in the Long Run

Thomas Piketty*, Gabriel Zucman†
*Paris School of Economics, Paris, France
†London School of Economics, London, UK

Contents

Handbook of Income Distribution, Volume 2B
ISSN 1574-0056, http://dx.doi.org/10.1016/B978-0-444-59429-7.00016-9
© 2015 Elsevier B.V.
All rights reserved.

Abstract

This chapter offers an overview of the empirical and theoretical research on the long-run evolution of wealth and inheritance. Wealth–income ratios, inherited wealth, and wealth inequalities were high in the eighteenth to nineteenth centuries up until World War I, then sharply dropped during the twentieth century following World War shocks, and have been rising again in the late twentieth and early twenty-first centuries. We discuss the models that can account for these facts. We show that over a wide range of models, the long-run magnitude and concentration of wealth and inheritance are an increasing function of $\bar{r} - g$ where $\bar{r}$ is the net-of-tax rate of return on wealth and g is the economy's growth rate. This suggests that current trends toward rising wealth–income ratios and wealth inequality might continue during the twenty-first century, both because of the slowdown of population and productivity growth, and because of rising international competition to attract capital.

Keywords

Wealth, Inheritance, Distribution, Growth, Rate of return, Pareto coefficient

JEL Classification Codes

E10, D30, D31, D32

15.1. INTRODUCTION

Economists have long recognized that the magnitude and distribution of wealth play an important role in the distribution of income—both across factors of production (labor and capital) and across individuals. In this chapter, we ask three simple questions: (1) What do we know about historical patterns in the magnitude of wealth and inheritance relative to income? (2) How does the distribution of wealth vary in the long run and across countries? (3) And what are the models that can account for these facts?

In surveying the literature on these issues, we will focus the analysis on three interrelated ratios. The first is the aggregate wealth-to-income ratio, that is the ratio between marketable—nonhuman—wealth and national income. The second is the share of aggregate wealth held by the richest individuals, say the top 10% or top 1%. The last is the ratio between the stock of inherited wealth and aggregate wealth (or between the annual flow of bequests and national income). As we shall see, to properly analyze the concentration of wealth and its implications, it is critical to study top wealth shares jointly with the macroeconomic wealth–income and inheritance–wealth ratios. In so doing, this chapter attempts to build bridges between income distribution and macroeconomics.

The wealth-to-income ratio, top wealth shares, and the share of inheritance in the economy have all been the subject of considerable interest and controversy—but usually on the basis of limited data. For a long time, economics textbooks have presented the wealth–income ratio as stable over time—one of the Kaldor facts.[1] There is, however, no strong theoretical reason why it should be so: With a flexible production function, any ratio can be a steady state. And until recently, we lacked comprehensive national balance sheets with harmonized definitions for wealth that could be used to vindicate the constant-ratio thesis. Recent research shows that wealth–income ratios, as well as the share of capital in national income, are actually much less stable in the long run than what is commonly assumed.

Following the Kuznets curve hypothesis, first formulated in the 1950s, another common view among economists has been that income inequality—and possibly wealth inequality as well—should first rise and then decline with economic development, as a growing fraction of the population joins high-productivity sectors and benefits from industrial growth.[2] However, following the rise in inequality that has occurred in most developed countries since the 1970s–1980s, this optimistic view has become less popular.[3] As a consequence, most economists are now fairly skeptical about universal laws regarding the long-run evolution of inequality.

Last, regarding the inheritance share in total wealth accumulation, there seems to exist a general presumption that it should tend to decline over time. Although this is rarely formulated explicitly, one possible mechanism could be the rise of human capital (leading maybe to a rise of the labor share in income and saving), or the rise in life-cycle wealth accumulation (itself possibly due to the rise of life expectancy). Until recently, however, there was limited empirical evidence on the share of inherited wealth available to test these hypotheses. The 1980s saw a famous controversy between Modigliani (a life-cycle advocate, who argued that the share of inherited wealth was as little as 20–30% of U.S. aggregate wealth) and Kotlikoff–Summers (who instead

[1] See, e.g., Kaldor (1961) and Jones and Romer (2010).
[2] See Kuznets (1953).
[3] See Atkinson et al. (2011). See also Chapter 7 in Handbook of Income Distribution, volume 2A by Roine and Waldenstrom (2015).

argued that the inheritance share was as large as 80%, if not larger). Particularly confusing was the fact that both sides claimed to look at the same data, namely U.S. data from the 1960s–1970s.[4]

Because many of the key predictions about wealth and inheritance were formulated a long time ago—often in the 1950s–1960s, or sometime in the 1970s–1980s—and usually on the basis of a relatively small amount of long-run evidence, it is high time to take a fresh look at them again on the basis of the more reliable evidence now available.

We begin by reviewing in Section 15.2 what we know about the historical evolution of the wealth–income ratio β. In most countries, this ratio has been following a U-shaped pattern over the 1910–2010 period, with a large decline between the 1910s and the 1950s, and a gradual recovery since the 1950s. The pattern is particularly spectacular in Europe, where the aggregate wealth–income ratio was as large as 600–700% during the eighteenth, nineteenth, and early twentieth centuries, then dropped to as little as 200–300% in the mid-twentieth century. It is now back to about 500–600% in the early twenty-first century. These same orders of magnitude also seem to apply to Japan, though the historical data is less complete than for Europe. The U-shaped pattern also exists—but is less marked—in the United States.

In Section 15.3, we turn to the long-run changes in wealth concentration. We also find a U-shaped pattern over the past century, but the dynamics have been quite different in Europe and in the United States. In Europe, the recent increase in wealth inequality appears to be more limited than the rise of the aggregate wealth–income ratio, so that European wealth seems to be significantly less concentrated in the early twenty-first century than a century ago. The top 10% wealth share used to be as large as 90%, whereas it is around 60–70% today (which is already quite large—and in particular a lot larger than the concentration of labor income). In the United States, by contrast, wealth concentration appears to have almost returned to its early twentieth century level. Although Europe was substantially more unequal than the United States until World War I, the situation has reversed over the course of the twentieth century. Whether the gap between both economies will keep widening in the twenty-first century is an open issue.

In Section 15.4, we describe the existing evidence regarding the evolution of the share φ of inherited wealth in aggregate wealth. This is an area in which available historical series are scarce and a lot of data has yet to be collected. However existing evidence—coming mostly from France, Germany, the United Kingdom, and Sweden—suggests that the inheritance share has also followed a U-shaped pattern over the past century. Modigliani's estimates—with a large majority of wealth coming from life-cycle savings—might have been right for the immediate postwar period (though somewhat exaggerated). But Kotlikoff–Summers' estimates—with inheritance

[4] See Kotlikoff and Summers (1981, 1988) and Modigliani (1986, 1988). Modigliani's theory of life-cycle saving was first formulated in the 1950s–1960s; see the references given in Modigliani (1986).

accounting for a significant majority of wealth—appear to be closer to what we generally observe in the long run, both in the nineteenth, twentieth, and early twenty-first centuries. Here again, there could be some interesting differences between Europe and the United States (possibly running in the opposite direction than for wealth concentration). Unfortunately the fragility of available U.S. data makes it difficult to conclude at this stage.

We then discuss in Section 15.5 the theoretical mechanisms that can be used to account for the historical evidence and to analyze future prospects. Some of the evolutions documented in Sections 15.2–15.4 are due to shocks. In particular, the large U-shaped pattern of wealth–income and inheritance-income ratios observed over the 1910–2010 period is largely due to the wars (which hit Europe and Japan much more than the United States). Here the main theoretical lesson is simply that capital accumulation takes time, and that the world wars of the twentieth century have had a long-lasting impact on basic economic ratios. This, in a way, is not too surprising and follows from simple arithmetic. With a 10% saving rate and a fixed income, it takes 50 years to accumulate the equivalent of 5 years of income in capital stock. With income growth, the recovery process takes even more time.

The more interesting and difficult part of the story is to understand the forces that determine the new steady-state levels toward which each economy tends to converge once it has recovered from shocks. In Section 15.5, we show that over a wide range of models, the long-run magnitude and concentration of wealth and inheritance are a decreasing function of g and an increasing function of $\bar{r}$, where g is the economy's growth rate and $\bar{r}$ is the net-of-tax rate of return to wealth. That is, under plausible assumptions, our three interrelated sets of ratios—the wealth–income ratio, the concentration of wealth, and the share of inherited wealth—all tend to take higher steady-state values when the long-run growth rate is lower or when the net-of-tax rate of return is higher. In particular, a higher $\bar{r} - g$ tends to magnify steady-state wealth inequalities. We argue that these theoretical predictions are broadly consistent with both the time-series and the cross-country evidence. This also suggests that the current trends toward rising wealth–income ratios and wealth inequality might continue during the twenty-first century, both because of population and productivity growth slowdown, and because of rising international competition to attract capital.

Owing to data availability constraints, the historical evolutions analyzed in this chapter relate for the most part to today's rich countries (Europe, North America, and Japan). However, to the extent that the theoretical mechanisms unveiled by the experience of rich countries also apply elsewhere, the findings presented here are also of interest for today's emerging economies. In Section 15.5, we discuss the prospects for the global evolution of wealth–income ratios, wealth concentration, and the share of inherited wealth in the coming decades. Finally, Section 15.6 offers concluding comments and stresses the need for more research in this area.

15.2. THE LONG-RUN EVOLUTION OF WEALTH–INCOME RATIOS

15.2.1 Concepts, Data Sources, and Methods

15.2.1.1 Country Balance Sheets

Prior to World War I, there was a vibrant tradition of national wealth accounting: economists, statisticians, and social arithmeticians were much more interested in computing the stock of national wealth than the flows of national income and output. The first national balance sheets were established in the late seventeenth and early eighteenth centuries by Petty (1664) and King (1696) in the United Kingdom, and Boisguillebert (1695) and Vauban (1707) in France. National wealth estimates then became plentiful in the nineteenth and early twentieth century, with the work of Colquhoun (1815), Giffen (1889), and Bowley (1920) in the United Kingdom, de Foville (1893) and Colson (1903) in France, Helfferich (1913) in Germany, King (1915) in the United States, and dozens of other economists.

The focus on wealth, however, largely disappeared in the interwar. The shock of World War I, the Great Depression, and the coming of Keynesian economics led to attention being switched from stocks to flows, with balance sheets being neglected. The first systematic attempt to collect historical balance sheets is due to Goldsmith (1985, 1991). Building upon recent progress made in the measurement of wealth, and pushing forward Goldsmith's pioneering attempt, Piketty and Zucman (2014) construct aggregate wealth and income series for the top eight rich economies. Other recent papers that look at specific countries include Atkinson (2013) for the United Kingdom and Ohlsson et al. (2013) for Sweden. In this section, we rely on the data collected by Piketty and Zucman (2014)—and closely follow the discussion therein—to present the long-run evolution of wealth–income ratios in the main developed economies.

In determining what is to be counted as wealth, we follow the U.N. System of National Accounts (SNA). For the 1970–2010 period, the data come from official national accounts that comply with the latest international guidelines (SNA, 1993, 2008). For the previous periods, Piketty and Zucman (2014) draw on the vast national wealth accounting tradition to construct homogenous income and wealth series that use the same concepts and definitions as in the most recent official accounts. The historical data themselves were established by a large number of scholars and statistical administrations using a wide variety of sources, including land, housing and wealth censuses, financial surveys, corporate book accounts, and the like. Although historical balance sheets are far from perfect, their methods are well documented and they are usually internally consistent. It was also somewhat easier to estimate national wealth around 1900–1910 than it is today: the structure of property was simpler, with less financial intermediation and cross-border positions.[5]

[5] A detailed analysis of conceptual and methodological issues regarding wealth measurement, as well as extensive country-specific references on historical balance sheets, are provided by Piketty and Zucman (2014).

15.2.1.2 Concepts and Definitions: Wealth Versus Capital

We define private wealth, W_t, as the net wealth (assets minus liabilities) of households.[6] Following SNA guidelines, assets include all the nonfinancial assets—land, buildings, machines, etc.—and financial assets—including life insurance and pensions funds—over which ownership rights can be enforced and that provide economic benefits to their owners. Pay-as-you-go Social Security pension wealth is excluded, just as all other claims on future government expenditures and transfers (such as education expenses for one's children or health benefits). Durable goods owned by households, such as cars and furniture, are excluded as well.[7] As a general rule, all assets and liabilities are valued at their prevailing market prices. Corporations are included in private wealth through the market value of equities and corporate bonds. Unquoted shares are typically valued on the basis of observed market prices for comparable, publicly traded companies.

Similarly, public (or government) wealth, W_{gt}, is the net wealth of public administrations and government agencies. In available balance sheets, public nonfinancial assets such as administrative buildings, schools, and hospitals are valued by cumulating past investment flows and upgrading them using observed real estate prices.

Market-value national wealth, W_{nt}, is the sum of private and public wealth:

$$W_{nt} = W_t + W_{gt}$$

and national wealth can also be decomposed into domestic capital and net foreign assets:

$$W_{nt} = K_t + NFA_t$$

In turn, domestic capital K_t can be written as the sum of agricultural land, housing, and other domestic capital (including the market value of corporations, and the value of other nonfinancial assets held by the private and public sectors, net of their liabilities).

Regarding income, the definitions and notations are standard. Note that we always use net-of-depreciation income and output concepts. National income Y_t is the sum of net domestic output and net foreign income: $Y_t = Y_{dt} + r_t \cdot NFA_t$.[8] Domestic output can be thought of as coming from some aggregate production function that uses domestic capital and labor as inputs: $Y_{dt} = F(K_t, L_t)$.

[6] Private wealth also includes the assets and liabilities held by nonprofit institutions serving households (NPISH). The main reason for doing so is that the frontier between individuals and private foundations is not always clear. In any case, the net wealth of NPISH is usually small, and always less than 10% of total net private wealth: currently it is about 1% in France, 3–4% in Japan, and 6–7% in the United States; see Piketty and Zucman (2014, Appendix Table A65). Note also that the household sector includes all unincorporated businesses.

[7] The value of durable goods appears to be relatively stable over time (about 30–50% of national income, i.e., 5–10% of net private wealth). See for instance Piketty and Zucman (2014, Appendix Table US.6f) for the long-run evolution of durable goods in the United States.

[8] National income also includes net foreign labor income and net foreign production taxes—both of which are usually negligible.

One might prefer to think about output as deriving from a two-sector production process (housing and nonhousing sectors), or more generally from n sectors. In the real world, the capital stock K_t comprises thousands of various assets valued at different prices (just like output Y_{dt} is defined as the sum of thousands of goods and services valued at different prices). We find it more natural, however, to start with a one-sector formulation. Since the same capital assets (i.e., buildings) are often used for housing and office space, it would be quite artificial to start by dividing capital and output into two parts. We will later on discuss the pros and cons of the one-sector model and the need to appeal to two-sector models and relative asset price movements to properly account for observed changes in the aggregate wealth–income ratio.

Another choice that needs to be discussed is the focus on market values for national wealth and capital. We see market values as a useful and well-defined starting point. But one might prefer to look at book values, for example, for short-run growth accounting exercises. Book values exceed market values when Tobin's Q is less than 1, and conversely when Tobin's Q is larger than 1. In the long run, however, the choice of book versus market value does not much affect the analysis (see Piketty and Zucman, 2014, for a detailed discussion).

We are interested in the evolution of the private wealth–national income ratio $\beta_t = W_t/Y_t$ and of the national wealth–national income ratio $\beta_{nt} = W_{nt}/Y_t$. In a closed economy, and more generally in an open economy with a zero net foreign position, the national wealth–national income ratio β_{nt} is the same as the domestic capital–output ratio $\beta_{kt} = K_t/Y_{dt}$.[9] If public wealth is equal to zero, then both ratios are also equal to the private wealth–national income ratio $\beta_t = \beta_{nt} = \beta_{kt}$. At the global level, the world wealth–income ratio is always equal to the world capital–output ratio.

15.2.2 The Very Long-Run: Britain and France, 1700–2010

Figures 15.1 and 15.2 present the very long-run evidence available for Britain and France regarding the national wealth–national income ratio β_{nt}. Net public wealth—either positive or negative—is usually a relatively small fraction of national wealth, so that the evolution of β_{nt} mostly reflects the evolution of the private wealth–national income ratio β_t (more on this below).[10]

[9] In principle, one can imagine a country with a zero net foreign asset position (so that $W_{nt} = K_t$) but nonzero net foreign income flows (so that $Y_t \neq Y_{dt}$). In this case the national wealth–national income ratio β_{nt} will slightly differ from the domestic capital–output ratio β_{kt}. In practice today, differences between Y_t and Y_{dt} are very small—national income Y_t is usually between 97% and 103% of domestic output Y_{dt} (see Piketty and Zucman, 2014, Appendix Figure A57). Net foreign asset positions are usually small as well, so that β_{kt} turns out to be usually close to β_{nt} in the 1970–2010 period (see Piketty and Zucman, 2014, Appendix Figure A67).

[10] For an historical account of the changing decomposition of national wealth into private and public wealth in Britain and France since the eighteenth century, see Piketty (2014, Chapter 3).

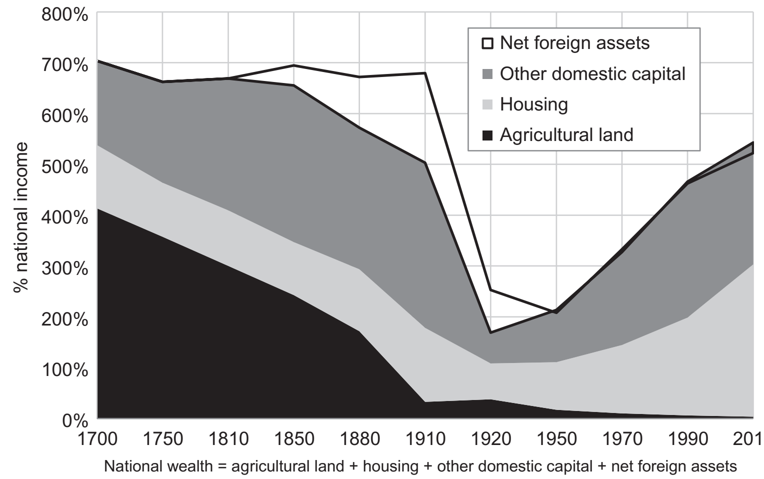


Figure 15.1 The changing level and nature of national wealth: United Kingdom 1700–2010.

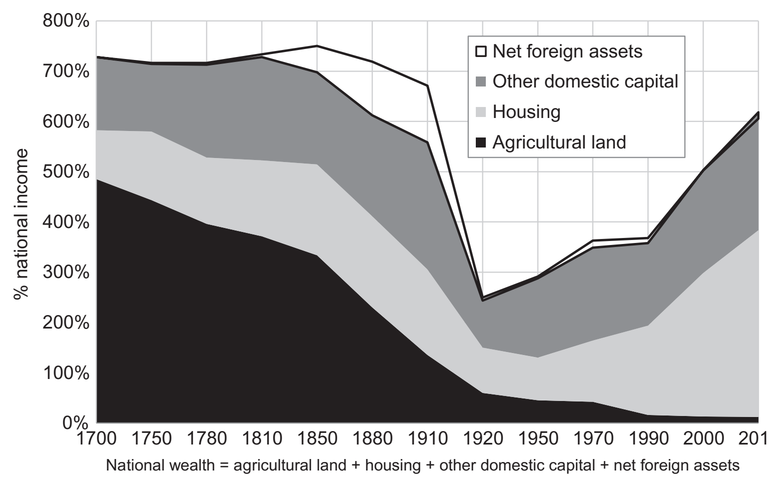


Figure 15.2 The changing level and nature of national wealth: France 1700–2010.

The evolutions are remarkably similar in the two countries. First, the wealth–income ratio has followed a spectacular U-shaped pattern. Aggregate wealth was worth about 6–7 years of national income during the eighteenth to nineteenth centuries on both sides of the channel, up until the eve of World War I. Raw data sources available for these two centuries are not sufficiently precise to make fine comparisons between the two countries or over time, but the orders of magnitude appear to be reliable and roughly stable (they come from a large number of independent estimates). Aggregate wealth then collapsed to as little as 2–3 years of national income in the aftermath of the two World Wars. Since the 1950s, there has been a gradual recovery in both countries. Aggregate wealth is back to about 5–6 years of national income in the 2000s to 2010s, just a bit below the pre-World War I level.

The other important finding that emerges from Figures 15.1 and 15.2 is that the composition of national wealth has changed in similar ways in both countries. Agricultural land, which made the majority of national capital in the eighteenth century, has been gradually replaced by real estate and other domestic capital, which is for the most part business capital (i.e., structures and equipment used by private firms). The nature of wealth has changed entirely reflecting a dramatic change in the structure of economic activity, and yet the total value of wealth is more or less the same as what it used to be before the Industrial Revolution.

Net foreign assets also made a large part of national capital in the late nineteenth century and on the eve of World War I: as much as 2 years of national capital in the case of Britain and over a year in the case of France. Net foreign-asset positions were brought back to zero in both countries following World War I and II shocks (including the loss of the colonial empires). In the late twentieth and early twenty-first centuries, net foreign positions are close to zero in both countries, just as in the eighteenth century. In the very long run, net foreign assets do not matter too much for the dynamics of the capital/income ratio in Britain or France. The main structural change is the replacement of agricultural land by housing and business capital.[11]

15.2.3 Old Europe Versus the New World

It is interesting to contrast the case of Old Europe—as illustrated by Britain and France—with that of the United States.

As Figure 15.3 shows, the aggregate value of wealth in the eighteenth to nineteenth centuries was markedly smaller in the New World than in Europe. At the time of the Declaration of Independence and in the early nineteenth century, national wealth in the United States was barely equal to 3–4 years of national income, about half that of Britain or France. Although available estimates are fragile, the order of magnitude again

[11] It is worth stressing that should we divide aggregate wealth by disposable household income (rather than national income), then today's ratios would be around 700–800% in Britain or France and would slightly surpass eighteenth to nineteenth century levels. This mechanically follows from the fact that disposable income was above 90% in the eighteenth to nineteenth centuries and is about 70–80% of disposable income in the late twentieth to early twenty-first century. The rising gap between disposable and household income reflects the rise of government-provided services, in particular in health and education. To the extent that these services are mostly useful (in their absence households would have to purchase them on the market), it is more justified for the purpose of historical and international comparisons to focus on ratios using national income as a denominator. For wealth–income ratios using disposable income as a denominator, see Piketty and Zucman (2014, Appendix, Figure A9).

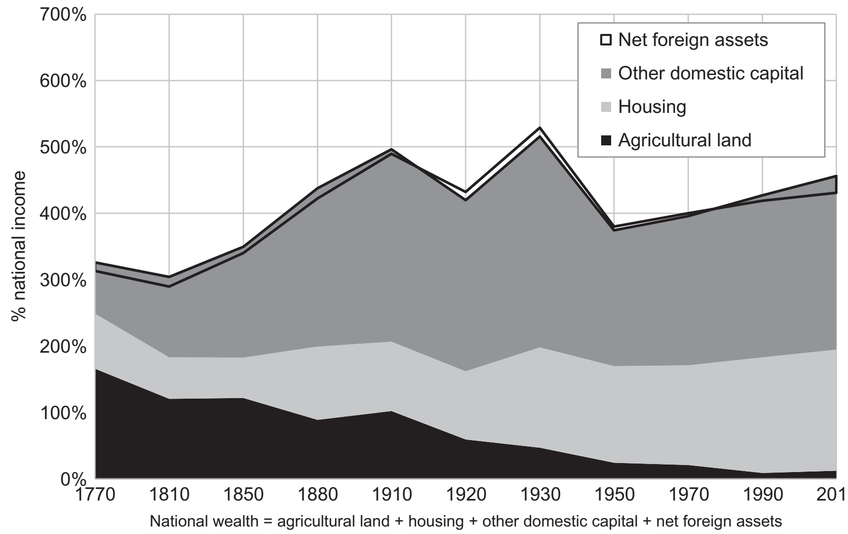


Figure 15.3 The changing level and nature of national wealth: United States 1770–2010.

appears to be robust. In Section 15.5, we will attempt to account for this interesting contrast. At this stage, we simply note that there are two obvious—and potentially complementary—factors that can play a role: first, there had been less time to save and accumulate wealth in the New World than in the Old World; second, there was so much land in the New World that it was almost worthless (its market value per acre was much less than in Europe).

The gap between the United States and Europe gradually reduces over the course of the nineteenth century, but still remains substantial. Around 1900–1910, national wealth is about 5 years of national income in the United States (see Figure 15.3) versus about 7 years in Britain and France. During the twentieth century, the U.S. wealth–income ratio also follows a U-shaped pattern, but less marked than in Europe. National wealth falls less sharply in the United States than in Europe following World War shocks, which seems rather intuitive. Interestingly, European wealth–income ratios have again surpassed U.S. ratios in the late twentieth and early twenty-first centuries.

This brief overview of wealth in the New World and Europe would be rather incomplete if we did not mention the issue of slavery. As one can see from Figure 15.4, the aggregate market value of slaves was fairly substantial in the United States until 1865: about 1–1.5 years of national income according to the best available historical sources. There were few slaves in Northern states, but in the South the value of the slave stock was so large that it approximately compensated—from the viewpoint of slave owners—the lower value of land as compared to the Old World (see Figure 15.5).

It is rather dubious, however, to include the market value of slaves into national capital. Slavery can be viewed as the most extreme form of debt: it should be counted as an

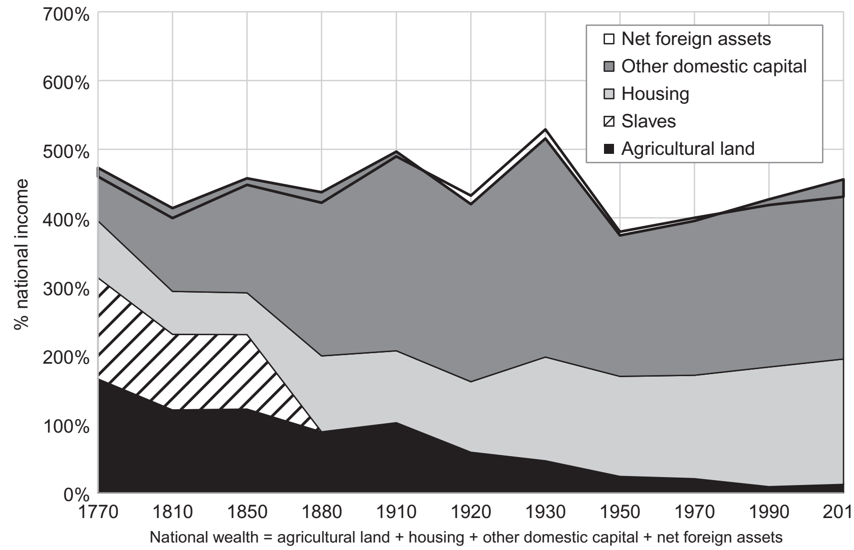


Figure 15.4 The changing level and nature of wealth: United States 1770–2010 (including slaves).

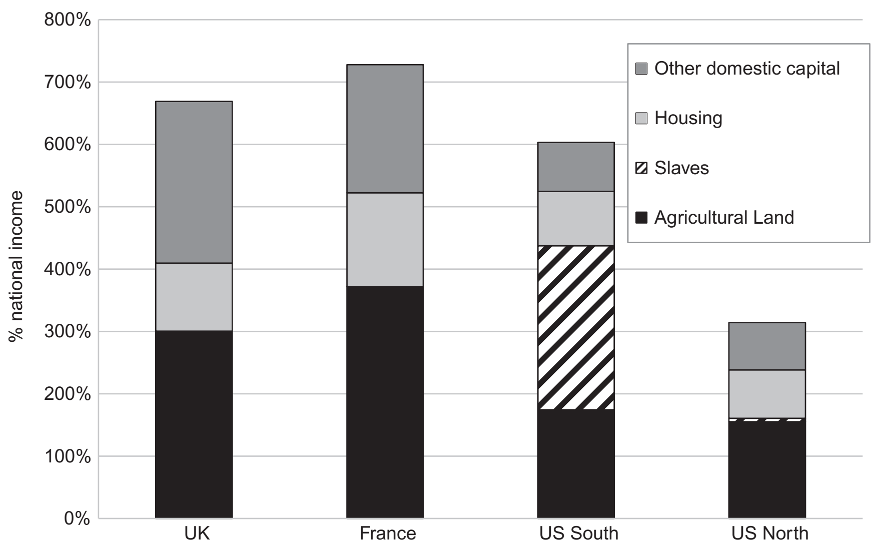


Figure 15.5 National wealth in 1770–1810: Old versus New World.

asset for the owners and a liability for the slaves, so that net national wealth should be unaffected. In the extreme case where a tiny elite owns the rest of the population, the total value of slaves—the total value of "human capital"—could be a lot larger than that of nonhuman capital (since the share of human labor in income is typically larger than

50%). If the rate of return r is equalized across all assets, then the aggregate value of human capital—expressed in proportion to national income—will be equal to $\beta_h = (1-\alpha)/r$, whereas the value of nonhuman capital will be given by $\beta_n = \alpha/r$, where α is the capital share and $1-\alpha$ the labor share implied by the production technology.[12] So for instance with $r = 5\%$, $\alpha = 30\%$, $1-\alpha = 70\%$, the value of the human capital stock will be as large as $\beta_h = (1-\alpha)/r = 1400\%$ (14 years of national income), and the value of the nonhuman capital stock will be $\beta_n = \alpha/r = 600\%$ (6 years of national income). Outside of slave societies, however, it is unclear whether it makes much sense to compute the market value of human capital and to add it to nonhuman capital.

The computations reported on Figures 15.4 and 15.5 illustrate the ambiguous relationship of the New World with wealth, inequality, and property. To some extent, America is the land of opportunity, the place where wealth accumulated in the past does not matter too much. But it is also the place where a new form of wealth and class structure—arguably more extreme and violent than the class structure prevailing in Europe—flourished, whereby part of the population owned another part.

Available historical series suggest that the sharp U-shaped pattern for the wealth–income ratio in Britain and France is fairly representative of Europe as a whole. For Germany, the wealth–income ratio was approximately the same as for Britain and France in the late nineteenth and early twentieth centuries, then fell to a very low level in the aftermath of the World Wars, and finally has been rising regularly since the 1950s (see Figure 15.6). Although the German wealth–income ratio is still below that of the United Kingdom and France, the speed of the recovery over the past few decades has been similar.[13] On Figure 15.7, we compare the European wealth–income ratio (obtained as a simple average of Britain, France, Germany, and Italy, the latter being available only for the most recent decades) to the U.S. one. The European wealth–income ratio was substantially above that of the United States until World War I, then fell significantly below in the aftermath of World War II, and surpassed it again in the late twentieth and early twenty-first centuries (see Figure 15.7).

[12] That is, $1-\alpha$ is the marginal product of labor times the labor (slave) stock. The formula $\beta_h = (1-\alpha)/r$ implicitly assumes that the fraction of output that is needed to feed and maintain the slave stock is negligible (otherwise it would just need to be deducted from $1-\alpha$), and that labor productivity is unaffected by the slavery condition (this is a controversial issue).

[13] The factors that can explain the lower German wealth–income ratio are the following. Real estate prices have increased far less in Germany than in Britain or France, which could be due in part to the lasting impact of German reunification and to stronger rent regulations. This could also be temporary. Next, the lower market value of German firms could be due to a stakeholder effect. Finally, the return to the German foreign portfolio, where a large part of German savings were directed, was particularly low in the most recent period. See Piketty and Zucman (2014, Section V.C) and Piketty (2014, Chapter 3).

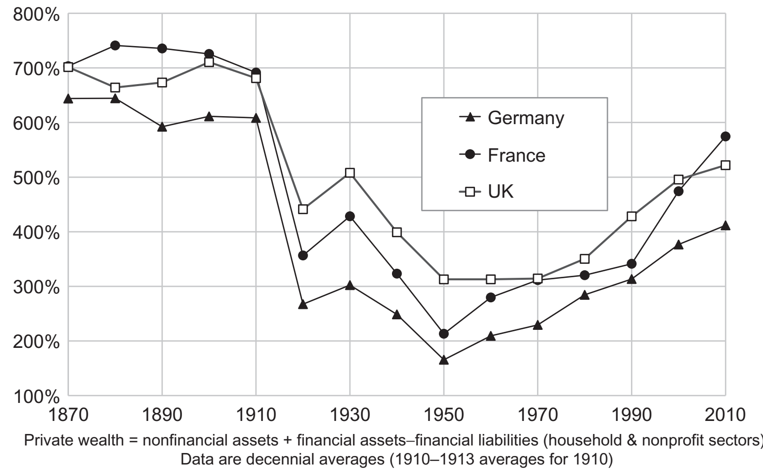


Figure 15.6 Private wealth/national income ratios in Europe, 1870–2010.

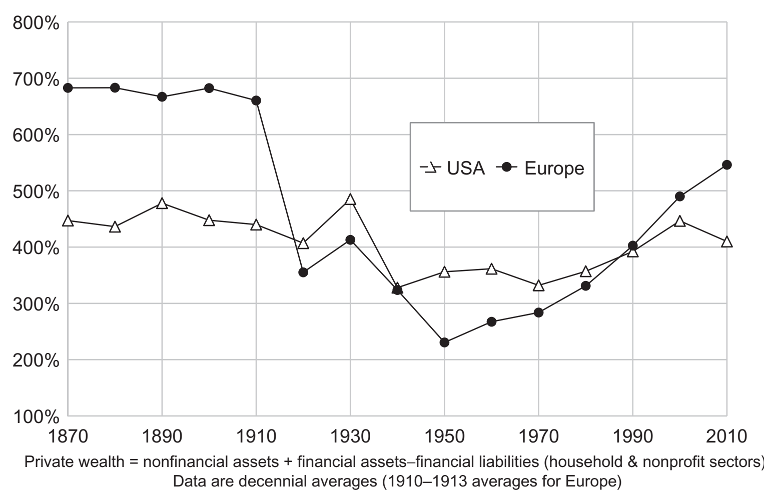


Figure 15.7 Private wealth/national income ratios 1870–2010: Europe versus United States.

15.2.4 The Return of High Wealth–Income Ratios in Rich Countries

Turning now to the 1970–2010 period, for which we have annual series covering most rich countries, the rise of wealth–income ratios, particularly private wealth–national income ratios, appears to be a general phenomenon. In the top eight developed economies, private wealth is between 2 and 3.5 years of national income around 1970, and between 4 and 7 years of national income around 2010 (see Figure 15.8). Although there are chaotic short-run fluctuations (reflecting the short-run volatility of asset prices),

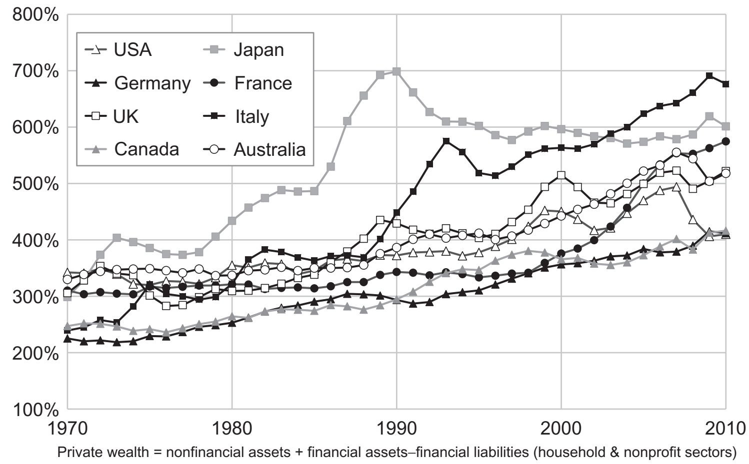


Figure 15.8 Private wealth/national income ratios, 1970–2010.

the long-run trend is clear. Take Japan. The huge asset price bubble of the late 1980s should not obscure the 1970–2010 rise of the wealth–income ratio, fairly comparable in magnitude to what we observe in Europe. (For instance, the Japanese and Italian patterns are relatively close: both countries go from about 2–3 years of national income in private wealth around 1970 to 6–7 years by 2010.)

Although we do not have national wealth estimates for Japan for the late nineteenth and early twentieth centuries, there are reasons to believe that the Japanese wealth–income ratio has also followed a U-shaped evolution in the long run, fairly similar to that observed in Europe over the twentieth century. That is, it seems likely that the wealth–income ratio was relatively high in the early twentieth century, fell to low levels in the aftermath of World War II, and then followed the recovery process that we see in Figure 15.8.[14]

To some extent, the rise of private wealth–national income ratios in rich countries since the 1970s is related to the decline of public wealth (see Figure 15.9). Public wealth has declined virtually everywhere owing both to the rise of public debt and the privatization of public assets. In some countries, such as Italy, public wealth has become strongly negative. The rise in private wealth, however, is quantitatively much larger than the decline in public wealth. As a result, national wealth—the sum of private and public wealth—has increased substantially, from 250–400% of national income in 1970 to 400–650% in 2010 (see Figure 15.10). In Italy, for instance, net government wealth fell by the equivalent of about 1 year of national income, but net private wealth rose by over

[14] The early twentieth century Japanese inheritance tax data reported by Morigushi and Saez (2008) are consistent with this interpretation.

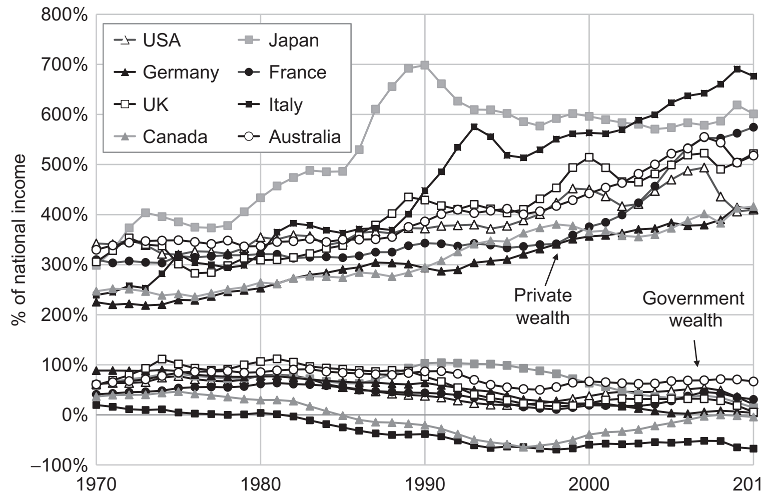


Figure 15.9 Private versus government wealth, 1970–2010.

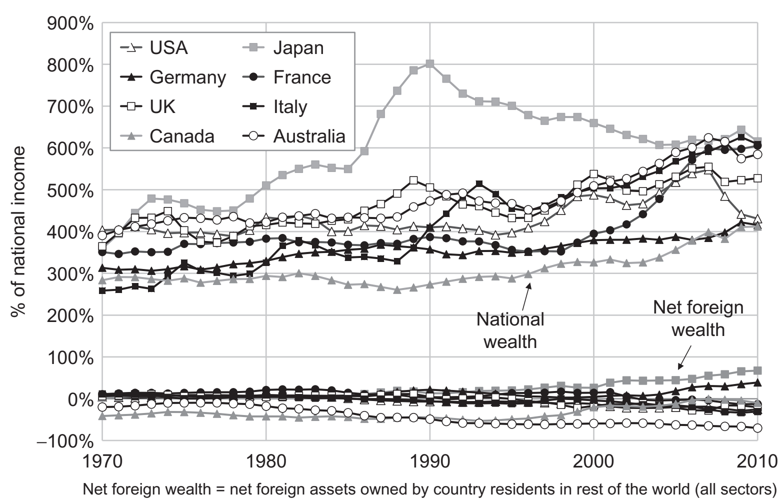


Figure 15.10 National versus foreign wealth, 1970–2010.

4 years of national income, so that national wealth increased by the equivalent of over 3 years of national income.

Figure 15.10 also depicts the evolution of net foreign wealth. Net foreign asset positions are generally small compared to national wealth. In other words, the evolution of national wealth–national income ratios mostly reflects the evolution of domestic capital–output ratios. There are two caveats, however. First, gross cross-border positions have risen a lot in recent decades, which can generate large portfolio valuation effects at the country level. Second, Japan and Germany have accumulated significant net foreign wealth (with net positions around 40% and 70% of national income, respectively, in 2010). Although these are still much smaller than the positions held by France and Britain on the eve of World War I (around 100% and 200% of national income, respectively), they are becoming relatively large (and were rising fast in the case of Germany in the first half of the 2010s, due to the large German trade surpluses).

15.3. THE LONG-RUN EVOLUTION OF WEALTH CONCENTRATION

15.3.1 Concepts, Data Sources, and Methods

We now turn to the evidence on the long-run evolution of wealth concentration. This question can be studied with different data sources (see Davies and Shorrocks, 1999, for a detailed discussion). Ideally, one would want to use annual wealth tax declarations for the entire population. Annual wealth taxes, however, often do to exist, and when they do, the data generally do not cover long periods of time.

The key source used to study the long-run evolution of wealth inequality has traditionally been inheritance and estate tax declarations.[15] By definition, estates and inheritance returns only provide information about wealth at death. The standard way to use inheritance tax data to study wealth concentration was invented over a century ago. Shortly before World War I, a number of British and French economists developed what is known as the mortality multiplier technique, whereby wealth-at-death is weighted by the inverse of the mortality rate of the given age and wealth group in order to generate estimates for the distribution of wealth among the living.[16] This approach was later followed in the United States by Lampman (1962) and Kopczuk and Saez (2004), who use estate tax data covering the 1916–1956 and 1916–2000 periods, respectively, and in the United Kingdom by Atkinson and Harrison (1978), who exploit inheritance tax data covering the 1922–1976 period.

To measure historical trends in the distribution of wealth, one can also use individual income tax returns and capitalize the dividends, interest, rents, and other forms of capital income declared on such returns. The capitalization technique was pioneered by King (1927), Stewart (1939), Atkinson and Harrison (1978), and Greenwood (1983), who used it to estimate the distribution of wealth in the United Kingdom and in the United States for some years in isolation. To obtain reliable results, it is critical to have detailed income data, preferably at the micro level, and to carefully reconcile the tax data with household balance sheets, so as to compute the correct capitalization factors. Drawing

[15] The difference between inheritance and estate taxes is that inheritance taxes are computed at the level of each inheritor, whereas estate taxes are computed at the level of the total estate (total wealth left by the decedent). The raw data coming from these two forms of taxes on wealth transfers are similar.

[16] See Mallet (1908), Séaillès (1910), Strutt (1910), Mallet and Strutt (1915), and Stamp (1919).

on the very detailed U.S. income tax data and Flow of Funds balance sheets, Saez and Zucman (2014) use the capitalization technique to estimate the distribution of U.S. wealth annually since 1913.

For the recent period, one can also use wealth surveys. Surveys, however, are never available on a long-run basis and raise serious difficulties regarding self-reporting biases, especially at the top of the distribution. Tax sources also raise difficulties at the top, especially for the recent period, given the large rise of offshore wealth (Zucman, 2013). Generally speaking, it is certainly more difficult for the recent period to accurately measure the concentration of wealth than the aggregate value of wealth, and one should be aware of this limitation. One needs to be pragmatic and combine the various available data sources (including the global wealth rankings published by magazines such as *Forbes*, which we will refer to in Section 15.5).

The historical series that we analyze in this chapter combines works by many different authors (more details below), who mostly relied on estate and inheritance tax data. They all relate to the inequality of wealth among the living.

We focus on simple concentration indicators, such as the share of aggregate wealth going to the top 10% individuals with the highest net wealth and the share going to the top 1%. In every country and historical period for which we have data, the share of aggregate wealth going to the bottom 50% is extremely small (usually less than 5%). So a decline in the top 10% wealth share can for the most part be interpreted as a rise in the share going to the middle 40%. Note also that wealth concentration is usually almost as large within each age group as for the population taken as a whole.[17]

15.3.2 The European Pattern: France, Britain, and Sweden, 1810–2010

15.3.2.1 France

We start with the case of France, the country for which the longest time series is available. French inheritance tax data is exceptionally good, for one simple reason. As early as 1791, shortly after the abolition of the tax privileges of the aristocracy, the French National Assembly introduced a universal inheritance tax, which has remained in force since then. This inheritance tax was universal because it applied both to bequests and to inter-vivos gifts, at any level of wealth, and for nearly all types of property (both tangible and financial assets). The key characteristic of the tax is that the successors of all decedents with positive wealth, as well as all donees receiving a positive gift, have always been required to file a return, no matter how small the estate was, and no matter whether any tax was ultimately owed.

In other countries, available data are less long run and/or less systematic. In the United Kingdom, one has to wait until 1894 for the unification of inheritance taxation (until this date the rules were different for personal and real estate taxes), and until the early 1920s

[17] See, e.g., Atkinson (1983) and Saez and Zucman (2014).

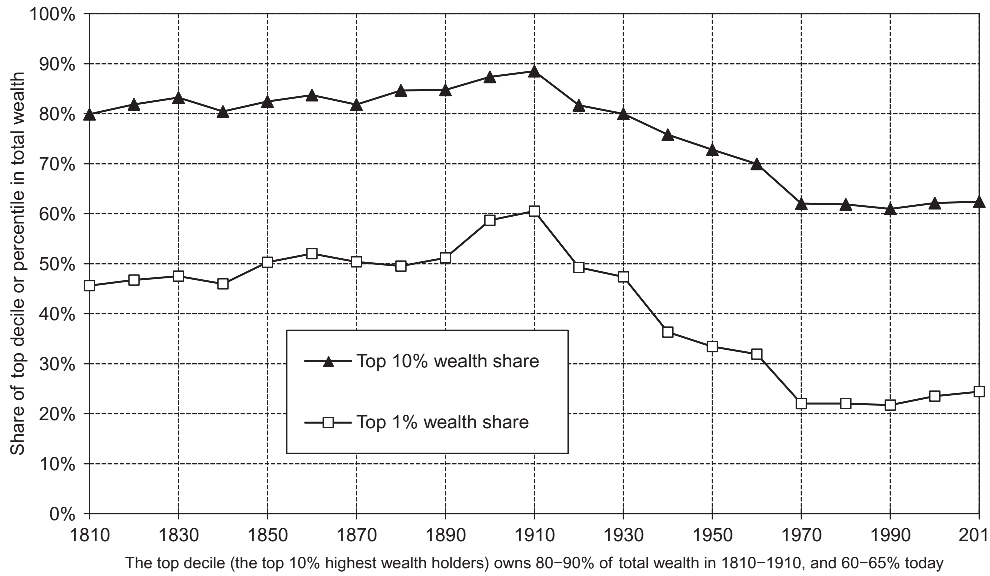


Figure 15.11 Wealth inequality in France, 1810–2010.

for unified statistics to be established by the U.K. tax administration. In the United States, one has to wait until 1916 for the creation of a federal estate tax and the publication of federal statistics on inheritance.

In addition, individual-level inheritance tax declarations have been well preserved in French national archives since the time of the revolution, so that one can use tax registers to collect large representative micro samples. Together with the tabulations by inheritance brackets published by the French tax administration, this allows for a consistent study of wealth inequality over a two-century-long period (see Piketty et al., 2006, 2013).

The main results are summarized on Figures 15.11 and 15.12.[18] First, wealth concentration was very high—and rising—in France during the nineteenth and early twentieth centuries. There was no decline in wealth concentration prior to World War I, quite the contrary: the trend toward rising wealth concentration did accelerate during the 1870–1913 period. The orders of magnitude are quite striking: in 1913, the top 10% wealth share is about 90%, and the top 1% share alone is around 60%. In Paris, which hosts about 5% of the population but as much as 25% of aggregate wealth, wealth is even more concentrated: more than two-thirds of the population has zero or negligible wealth, and 1% of the population owns 70% of the wealth.

[18] The updated series used for Figures 15.11 and 15.12 are based on the historical estimates presented by Piketty et al. (2006) and more recent fiscal data. See Piketty (2014, Chapter 10, Figures 10.1–10.2).

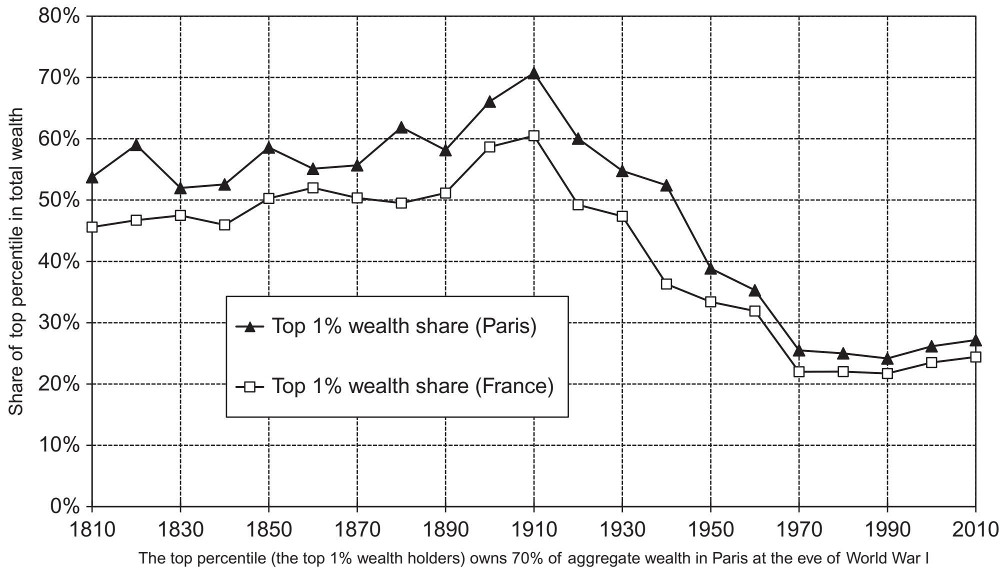


Figure 15.12 Wealth inequality: Paris versus France, 1810–2010.

Looking at Figures 15.11 and 15.12, one naturally wonders whether wealth concentration would have kept increasing without the 1914–1945 shocks. It might have stabilized at a very high level, but it could also have started to decline at some point. In any case, it is clear that the war shocks induced a violent regime change.

The other interesting fact is that wealth concentration has started to increase again in France since the 1970s–1980s—but it is still much lower than on the eve of World War I. According to the most recent data, the top 10% wealth share is slightly above 60%. Given the relatively low quality of today's wealth data, especially regarding top global wealth holders, one should be cautious about this estimate. It could well be that we somewhat underestimate the recent rise and the current level of wealth concentration.[19] In any case, a share of 60% for the top decile is already high, especially compared to the concentration of labor income: the top 10% of labor earners typically receive less than 30% of aggregate labor income.

15.3.2.2 Britain

Although the data sources for other countries are not as systematic and comprehensive as the French sources, existing evidence suggests that the French pattern extends to other European countries. For the United Kingdom, on Figure 15.13, we have combined historical estimates provided by various authors—particularly Atkinson and Harrison (1978)

[19] In contrast, the nineteenth and early twentieth centuries estimates are probably more precise (the tax rates were so low at that time that there was little incentive to hide wealth).

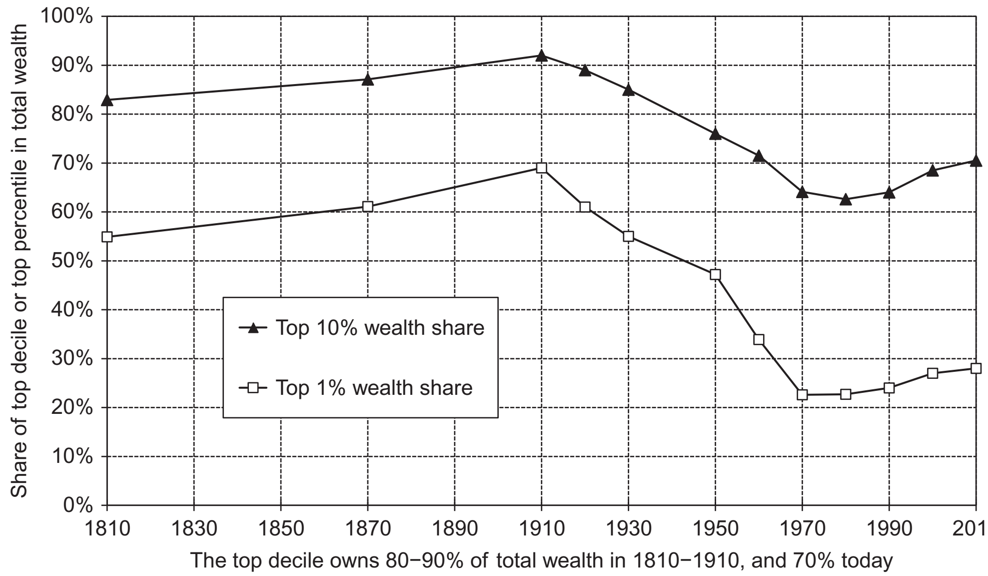


Figure 15.13 Wealth inequality in the United Kingdom, 1810–2010.

and Lindert (1986)—as well as more recent estimates using inheritance tax data. These series are not fully homogenous (in particular, the nineteenth century computations are based on samples of private probate records and are not entirely comparable to the twentieth-century inheritance tax data), but they deliver a consistent picture. Wealth concentration was high and rising during the nineteenth century up until World War I, then fell abruptly following the 1914–1945 shocks, and has been rising again since the 1980s.

According to these estimates, wealth concentration was also somewhat larger in the United Kingdom than in France in the nineteenth and early twentieth centuries. Yet the gap is much smaller than what French contemporary observers claimed. Around 1880–1910, it was very common among French republican elites to describe France as a "country of little property owners" (*un pays de petits propriétaires*), in contrast to aristocratic Britain. Therefore, the argument goes, there was no need to introduce progressive taxation in France (this should be left to Britain). The data show that on the eve of World War I the concentration of wealth was almost as extreme on both sides of the channel: the top 10% owns about 90% of wealth in both countries, and the top 1% owns 70% of wealth in Britain, versus 60% in France. It is true that aristocratic landed estates were more present in the United Kingdom (and to some extent still are today). But given that the share of agricultural land in national wealth dropped to low levels during the nineteenth century (see Figures 15.1 and 15.2), this does not matter much. At the end of the day, whether the country is a republic or a monarchy seems to have little impact on wealth concentration in the long run.

15.3.2.3 Sweden

Although widely regarded as an egalitarian haven today, Sweden was just as unequal as France and Britain in the nineteenth and early twentieth centuries. This is illustrated by Figure 15.13, where we plot some of the estimates constructed by Roine and Waldenstrom (2009) and Waldenstrom (2009).

The concentration of wealth is quite similar across European countries, both for the more ancient and the more recent estimates. Beyond national specificities, a European pattern emerges: the top 10% wealth share went from about 90% around 1900–1910 to about 60–70% in 2000–2010, with a recent rebound. In other words, about 20–30% of national wealth has been redistributed away from the top 10% to the bottom 90%. Since most of this redistribution benefited the middle 40% (the bottom 50% still hardly owns any wealth), this evolution can be described as the rise of a patrimonial middle class (Figure 15.14).

In the case of Sweden, Roine and Waldenstrom (2009) have also computed a corrected top 1% of wealth shares using estimates of offshore wealth held abroad by rich Swedes. They find that under plausible assumptions the top 1% share would shift from about 20% of aggregate wealth to over 30% (i.e., approximately the levels observed in the United Kingdom, and not too far away from the level observed in the United States). This illustrates the limitations of our ability to measure recent trends and levels, given the rising importance of tax havens.

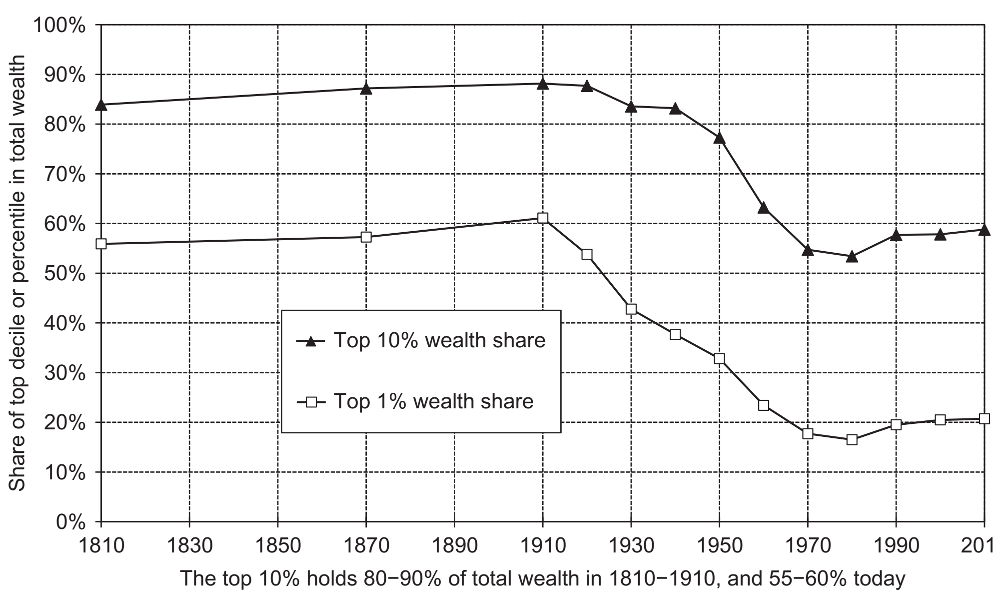


Figure 15.14 Wealth inequality in Sweden, 1810–2010.

15.3.3 The Great Inequality Reversal: Europe Versus the United States, 1810–2010

Comparing wealth concentration in Europe and the United States, the main finding is a fairly spectacular reversal. In the nineteenth century, the United States was to some extent the land of equality (at least for white men): the concentration of wealth was much less extreme than in Europe (except in the South). Over the course of the twentieth century, this ordering was reversed: wealth concentration has become significantly higher in the United States. This is illustrated by Figure 15.15, where we combine the estimates due to Lindert (2000) for the nineteenth century with those of Saez and Zucman (2014) for the twentieth and twenty-first centuries to form long-run U.S. series, and by Figure 15.16, where we compare the United States to Europe (defined as the arithmetic average of France, Britain, and Sweden).

The reversal comes from the fact that Europe has become significantly less unequal over the course of the twentieth century, whereas the United States has not. The United States has almost returned to its early twentieth-century wealth concentration level: at its peak in the late 1920s, the 10% wealth share was about 80%, in 2012 it is about 75%; similarly the top 1% share peaked at about 45% and is back to around 40% today. Note, however, that the United States never reached the extreme level of wealth concentration of nineteenth- and early twentieth-century Europe (with a top decile of 90% or more). The United States has always had a patrimonial middle class, although one of varying importance. The share of wealth held by the middle class appears to have been shrinking since the 1980s.

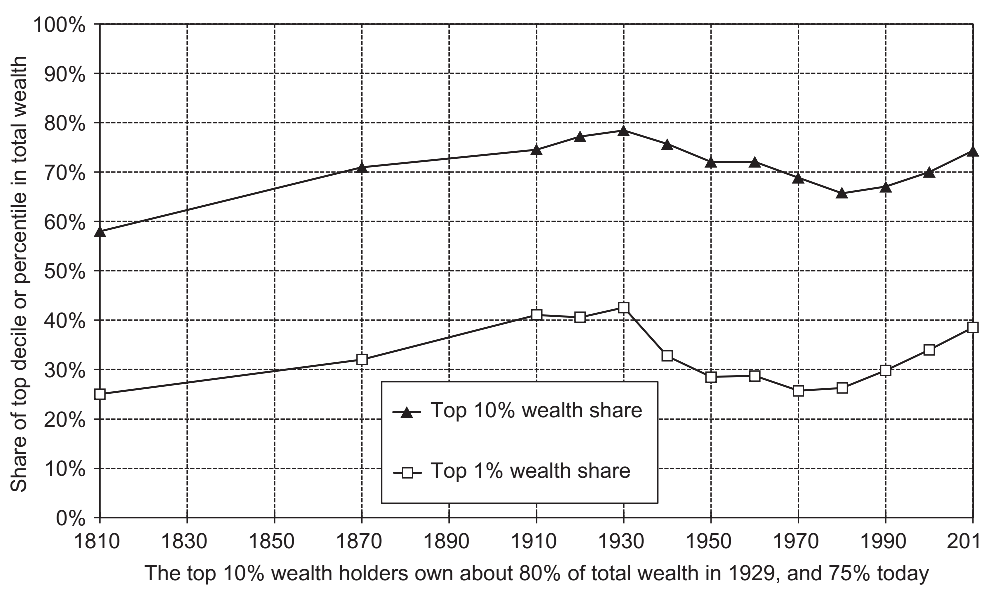


Figure 15.15 Wealth inequality in the United States, 1810–2010.

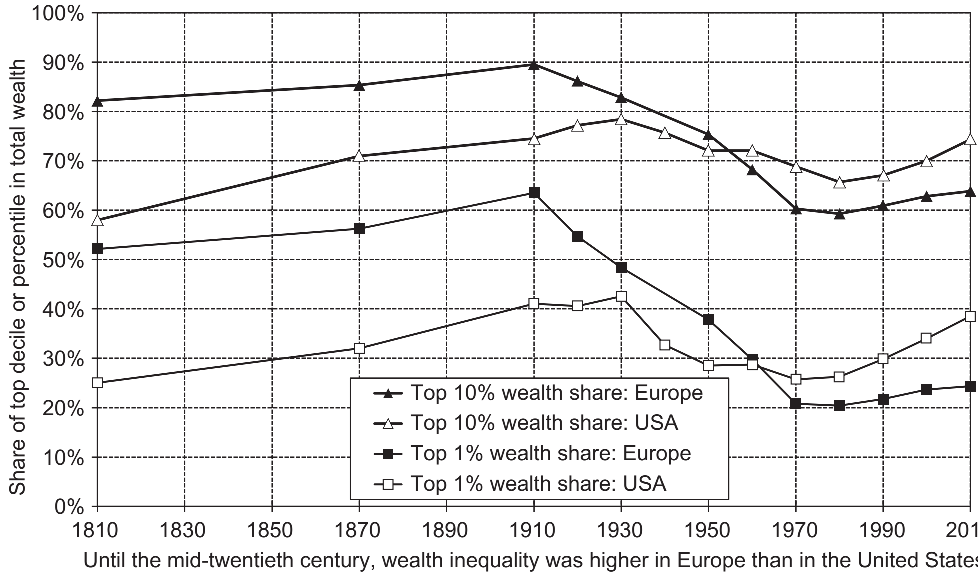


Figure 15.16 Wealth inequality: Europe and the United States, 1810–2010.

U.S. economists of the early twentieth century were very concerned about the possibility that their country becomes as unequal as Old Europe. Irving Fisher, then president of the American Economic Association, gave his presidential address in 1919 on this topic. He argued that the concentration of income and wealth was becoming as dangerously excessive in America as it had been for a long time in Europe. He called for steep tax progressivity to counteract this tendency. Fisher was particularly concerned about the fact that as much as half of U.S. wealth was owned by just 2% of U.S. population, a situation that he viewed as "undemocratic" (see Fisher, 1920). One can indeed interpret the spectacular rise of tax progressivity that occurred in the United States during the first half of the twentieth century as an attempt to preserve the egalitarian, democratic American ethos (celebrated a century before by Tocqueville and others). Attitudes toward inequality are dramatically different today. Many U.S. observers now view Europe as excessively egalitarian (and many European observers view the United States as excessively nonegalitarian).

15.4. THE LONG-RUN EVOLUTION OF THE SHARE OF INHERITED WEALTH

15.4.1 Concepts, Data Sources, and Methods

We now turn to our third ratio of interest, the share of inherited wealth in aggregate wealth. We should make clear at the outset that this is an area where available evidence is scarce and incomplete. Measuring the share of inherited wealth requires a lot more data than the measurement of aggregate wealth–income ratios or even wealth concentration. It is also an area where it is important to be particularly careful about concepts and definitions. Purely definitional conflicts have caused substantial confusion in the past. Therefore it is critical to start from there.

15.4.1.1 Basic Notions and Definitions

The most natural way to define the share of inherited wealth in aggregate wealth is to cumulate past inheritance flows. That is, assume that we observe the aggregate wealth stock W_t at time t in a given country, and that we would like to define and estimate the aggregate inherited wealth stock $W_{Bt} \leq W_t$ (and conversely aggregate self-made wealth, which we simply define as $W_{St} = W_t - W_{Bt}$). Assume that we observe the annual flow of inheritance B_s that occurred in any year $s \leq t$. At first sight, it might seem natural to define the stock of inherited wealth W_{Bt} as the sum of past inheritance flows:

$$W_{Bt} = \int_{s \leq t} B_s \cdot \mathrm{d}s$$

However, there are several practical and conceptual difficulties with this ambiguous definition, which need to be addressed before the formula can be applied to actual data. First, it is critical to include in this sum not only past bequest flows B_s (wealth transmissions at death) but also inter vivos gift flows V_s (wealth transmissions inter vivos). That is, one should define W_{Bt} as $W_{Bt} = \int_{s \leq t} B_s^* \cdot \mathrm{d}s$, with $B_s^* = B_s + V_s$.

Alternatively, if one cannot observe directly the gift flow V_s, one should replace the observed bequest flow B_s by some gross level $B_s^* = (1 + v_s) \cdot B_s$, where $v_s = V_s / B_s$ is an estimate of the gift/bequest flow ratio. In countries where adequate data is available, the gift–bequest ratio is at least 10–20%, and is often higher than 50%, especially in the recent period.[20] It is thus critical to include gifts in one way or another. In countries where fiscal data on gifts are insufficient, one should at least try to estimate $1 + v_s$ using surveys (which often suffers from severe downward biases) and harder administrative evidence from other countries.

Next, to properly apply this definition, one should only take into account the fraction of the aggregate inheritance flow $B_{st} \leq B_s$ that was received at time s by individuals who are still alive at time t. The problem is that doing so properly requires very detailed individual-level information. At any time t, there are always individuals who received inheritance a very long time ago (say, 60 years ago) but who are still alive (because they inherited at a very young age and/or are enjoying a very long life). Conversely, a fraction

[20] See below. Usually one only includes formal, monetary capital gifts, and one ignores informal presents and in-kind gifts. In particular in-kind gifts made to minors living with their parents (i.e., the fact that minor children are usually catered by their parents) are generally left aside.

of the inheritance flow received a short time ago (say, 10 years ago) should not be counted (because the relevant inheritors are already dead, e.g., they inherited at an old age or died young). In practice, however, such unusual events tend to balance each other, so that a standard simplifying assumption is to cumulate the full inheritance flows observed the previous H years, where H is the average generation length, that is, the average age at which parents have children (typically $H=30$ years). Therefore we obtain the following simplified definition:

$$W_{Bt} = \int_{t-30 \le s \le t} (1+v_s) \cdot B_s \cdot \mathrm{d}s$$

15.4.1.2 The Kotlikoff–Summers–Modigliani Controversy

Assume now that these two difficulties can be addressed (i.e., that we can properly estimate the factor $1+v_s$ and the average generation length H). There are more substantial difficulties ahead. First, to properly compute W_{Bt}, one needs to be able to observe inheritance flows B_s^* over a relatively long time period (typically, the previous 30 years). In the famous Kotlikoff–Summers–Modigliani (KSM) controversy, both Kotlikoff and Summers (1981) and Modigliani (1986, 1988) used estimates of the U.S. inheritance flow for only 1 year (and a relatively ancient year: 1962), see also Kotlikoff (1988). They simply assumed that this estimate could be used for other years. Namely, they assumed that the inheritance flow–national income ratio (which we note $b_{ys} = B_s^*/Y_s$) is stable over time. One problem with this assumption is that it might not be verified. As we shall see below, extensive historical data on inheritances recently collected in France show that the b_{ys} ratio has changed tremendously over the past two centuries, from about 20–25% of national income in the nineteenth and early twentieth centuries, down to less than 5% at mid-twentieth century, back to about 15% in the early twenty-first century (Piketty, 2011). So one cannot simply use one year of data and assume that we are in a steady state: One needs a long-run time series on the inheritance flow in order to estimate the aggregate stock of inherited wealth.

Next, one needs to decide the extent to which past inheritance flows need to be upgraded or capitalized. This is the main source of disagreement and confusion in the KSM controversy.

Modigliani (1986, 1988) chooses zero capitalization. That is, he simply defines the stock of inherited wealth W_{Bt}^M as the raw sum of past inheritance flows with no adjustment whatsoever (except for the GDP price index):

$$W_{Bt}^M = \int_{t-30 \le s \le t} B_s^* \cdot \mathrm{d}s$$

Assume a fixed inheritance flow–national income ratio $b_y = B_s^*/Y_s$, growth rate g (so that $Y_t = Y_s \cdot e^{g(t-s)}$), generation length H, and aggregate private wealth–national income ratio

$\beta = W_t/Y_t$. Then, according to the Modigliani definition, the steady-state formulas for the stock of inherited wealth relative to national income W_{Bt}^M/Y_t and for the share of inherited wealth $\varphi_t^M = W_{Bt}^M/W_t$ are given by

$$W_{Bt}^M/Y_t = \frac{1}{Y_t}\int_{t-30\leq s\leq t} B_s^* \cdot \mathrm{d}s = \frac{1-\mathrm{e}^{-gH}}{g}\cdot b_y$$

$$\varphi_t^M = W_{Bt}^M/W_t = \frac{1-\mathrm{e}^{-gH}}{g}\cdot\frac{b_y}{\beta}$$

In contrast, Kotlikoff and Summers (1981, 1988) choose to capitalize past inheritance flows by using the economy's average rate of return to wealth (assuming it is constant and equal to r). Following the Kotlikoff–Summers definition, the steady-state formulas for the stock of inherited wealth relative to national income W_{Bt}^{KS}/Y_t and for the share of inherited wealth $\varphi_t^{KS} = W_{Bt}^{KS}/W_t$ are given by

$$W_{Bt}^{KS}/Y_t = \frac{1}{Y_t}\int_{t-30\leq s\leq t} \mathrm{e}^{r(t-s)}\cdot B_s^* \cdot \mathrm{d}s = \frac{\mathrm{e}^{(r-g)H}-1}{r-g}\cdot b_y$$

$$\varphi_t^{KS} = W_{Bt}^{KS}/W_t = \frac{\mathrm{e}^{(r-g)H}-1}{r-g}\cdot\frac{b_y}{\beta}$$

In the special case where growth rates and rates of return are negligible (i.e., infinitely close to zero), then both definitions coincide. That is, if $g=0$ and $r-g=0$, then $(1-\mathrm{e}^{-gH})/g = (\mathrm{e}^{(r-g)H}-1)/(r-g) = H$, so that $W_{Bt}^M/Y_t = W_{Bt}^{KS}/Y_t = Hb_y$ and $\varphi_t^M = \varphi_t^{KS} = Hb_y/\beta$.

Thus, in case growth and capitalization effects can be neglected, one simply needs to multiply the annual inheritance flow by generation length. If the annual inheritance flow is equal to $b_y = 10\%$ of national income, and generation length is equal to $H = 30$ years, then the stock of inherited wealth is equal to $W_{Bt}^M = W_{Bt}^{KS} = 300\%$ of national income according to both definitions. In case aggregate wealth amounts to $\beta = 400\%$ of national income, then the inheritance share is equal to $\varphi_t^M = \varphi_t^{KS} = 75\%$ of aggregate wealth.

However, in the general case where g and $r-g$ are significantly different from zero, the two definitions can lead to widely different conclusions. For instance, with $g=2\%$, $r=4\%$, and $H=30$, we have the following capitalization factors: $(1-\mathrm{e}^{-gH})/(g\cdot H) = 0.75$ and $(\mathrm{e}^{(r-g)H}-1)/((r-g)\cdot H) = 1.37$. In this example, for a given inheritance flow $b_y = 10\%$ and aggregate wealth–income ratio $\beta = 400\%$, we obtain $\varphi_t^M = 56\%$ and $\varphi_t^{KS} = 103\%$. About half of wealth comes from inheritance according to the Modigliani definition, and all of it according to the Kotlikoff–Summers definition.

This is the main reason why Modigliani and Kotlikoff–Summers disagree so much about the inheritance share. They both use the same (relatively fragile) estimate for the United States b_y in 1962. But Modigliani does not capitalize past inheritance flows and concludes that the inheritance share is as low as 20–30%. Kotlikoff–Summers do capitalize the same flows and conclude that the inheritance share is as large as 80–90% (or even larger than 100%). Both sides also disagree somewhat about the measurement of b_y, but the main source of disagreement comes from this capitalization effect.[21]

15.4.1.3 The Limitations of KSM Definitions

Which of the two definitions is most justified? In our view, both are problematic. It is wholly inappropriate not to capitalize at all past inheritance flows. But full capitalization is also inadequate.

The key problem with the KSM representative-agent approach is that it fails to recognize that the wealth accumulation process always involves two different kinds of people and wealth trajectories. In every economy, there are inheritors (people who typically consume part of the return to their inherited wealth) and there are savers (people who do not inherit much but do accumulate wealth through labor income savings). This is an important feature of the real world that must be taken into account for a proper understanding of the aggregate wealth accumulation process.

The Modigliani definition is particularly problematic as it simply fails to recognize that inherited wealth produces flow returns. This mechanically leads to artificially low numbers for the inheritance share φ_t^M (as low as 20–40%), and to artificially high numbers for the life-cycle share in wealth accumulation, which Modigliani defines as $1-\varphi_t^M$ (up to 60–80%). As Blinder (1988) argues, "a Rockefeller with zero lifetime labor income and consuming only part of his inherited wealth income would appear to be a life-cycle saver in Modigliani's definition, which seems weird to me." One can easily construct illustrative examples of economies where all wealth comes from inheritance (with dynasties of the sort described by Blinder), but where Modigliani would still find an inheritance share well below 50%, simply because of his definition. This makes little sense.[22]

The Kotlikoff–Summers definition is conceptually more satisfactory than Modigliani's. But it suffers from the opposite drawback in the sense that it mechanically leads to artificially high numbers for the inheritance share φ_t^{KS}. In particular, φ_t^{KS} can easily be larger than 100%, even though there are life-cycle savers and self-made wealth accumulators in the economy, and a significant fraction of aggregate wealth accumulation comes from them. This will arise whenever the cumulated return to inherited wealth

[21] In effect, Modigliani favors a b_y ratio around 5–6%, whereas Kotlikoff–Summers find it more realistic to use a b_y ratio around 7–8%. Given the data sources they use, it is likely that both sides tend to somewhat underestimate the true ratio. See the discussion below for the case of France and other European countries.

[22] It is worth stressing that the return to inherited wealth (and the possibility to save and accumulate more wealth out of the return to inherited wealth) is a highly relevant economic issue not only for high-wealth dynasties of the sort referred to by Blinder, but also for middle-wealth dynasties. For instance, it is easier to save if one has inherited a house and has no rent to pay. An inheritor saving less than the rental value of his inherited home would be described as a life-cycle saver according to Modigliani's definition, which again seems odd.

consumed by inheritors exceeds the savers' wealth accumulation from their labor savings. In the real world, this condition seems to hold not only in prototype rentier societies such as Paris 1872–1937 (see Piketty et al., 2013), but also in countries and time periods when aggregate inheritance flow is relatively low. For instance, aggregate French series show that the capitalized bequest share φ_t^{KS} has been larger than 100% throughout the twentieth century, including in the 1950s–1970s, a period where a very significant amount of new self-made wealth was accumulated (Piketty, 2011).

In sum, the Modigliani definition leads to estimates of the inheritance share that are artificially close to 0%, whereas the Kotlikoff–Summers leads to inheritance shares that tend to be structurally above 100%. Neither of them offers an adequate way to look at the data.

15.4.1.4 The PPVR Definition

In an ideal world with perfect data, the conceptually consistent way to define the share of inherited wealth in aggregate wealth is the following. It has first been formalized and applied to Parisian wealth data by Piketty et al. (2013), so we refer to it as the PPVR definition.

The basic idea is to split the population into two groups. First, there are "inheritors" (or "rentiers"), whose assets are worth less than the capitalized value of the wealth they inherited (over time they consume more than their labor income). The second group is composed of "savers" (or "self-made individuals"), whose assets are worth more than the capitalized value of the wealth they inherited (they consume less than their labor income). Aggregate inherited wealth can then be defined as the sum of inheritors' wealth plus the inherited fraction of savers' wealth, and self-made wealth as the noninherited fraction of savers' wealth. By construction, inherited and self-made wealth are less than 100% and sum to aggregate wealth, which is certainly a desirable property. Although the definition is fairly straightforward, it differs considerably from the standard KSM definitions based on representative agent models. The PPVR definition is conceptually more consistent and provides a more meaningful way to look at the data and to analyze the structure of wealth accumulation processes. In effect, it amounts to defining inherited wealth at the individual level as the minimum between current wealth and capitalized inheritance.

More precisely, consider an economy with population N_t at time t. Take a given individual i with wealth w_{ti} at time t. Assume he or she received bequest b_{ti}^0 at time $t_i < t$. Note $b_{ti}^* = b_{ti}^0 \cdot e^{r(t-t_i)}$ the capitalized value of b_{ti}^0 at time t (where $e^{r(t-ti)}$ is the cumulated rate of return between time t_i and time t). Individual i is said to be an "inheritor" (or a "rentier") if $w_{ti} < b_{ti}^*$ and a "saver" (or a "self-made individual") if $w_{ti} \geq b_{ti}^*$. We define the set of inheritors as $N_t^r = \{i \,\text{s.t.}\, w_{ti} < b_{ti}^*\}$ and the set of savers as $N_t^s = \{i \,\text{s.t.}\, w_{ti} \geq b_{ti}^*\}$.

We note $\rho_t = N_t^r / N_t$ and $1 - \rho_t = N_t^s / N_t$ as the corresponding population shares of inheritors and savers; $w_t^r = E(w_{ti} | w_{ti} < b_{ti}^*)$ and $w_t^s = E(w_{ti} | w_{ti} \geq b_{ti}^*)$ as the average wealth levels of both groups; $br_t^* = E(b_{ti}^* | w_{ti} < b_{ti}^*)$ and $bs_t^* = E(b_{ti}^* | w_{ti} \geq b_{ti}^*)$ as the levels of their average capitalized bequest; and $\pi_t = \rho_t \cdot w_t^r / w_t$ and $1 - \pi_t = (1 - \rho_t) \cdot w_t^s / w_t$ as the share of inheritors and savers in aggregate wealth.

We define the total share φ_t of inherited wealth in aggregate wealth as the sum of inheritors' wealth plus the inherited fraction of savers' wealth, and the share $1 - \varphi_t$ of self-made wealth as the noninherited fraction of savers' wealth:

$$\varphi_t = \left[\rho_t \cdot w_t^r + (1 - \rho_t) \cdot b_t^{s*}\right] / w_t = \pi_t + (1 - \rho_t) \cdot b_t^{s*} / w_t$$

$$1 - \varphi_t = (1 - \rho_t) \cdot \left(w_t^s - b_t^{s*}\right) / w_t = 1 - \pi_t - (1 - \rho_t) \cdot b_t^{s*} / w_t$$

The downside of this definition is that it is more demanding in terms of data availability. Although Modigliani and Kotlikoff–Summers could compute inheritance shares in aggregate wealth by using aggregate data only, the PPVR definition requires micro data. Namely, we need data on the joint distribution $G_t(w_{ti}, b_{ti}^*)$ of current wealth w_{ti} and capitalized inherited wealth b_{ti}^* in order to compute ρ_t, π_t, and φ_t. This does require high-quality, individual-level data on wealth and inheritance over two generations, which is often difficult to obtain. It is worth stressing, however, that we do not need to know anything about the individual labor income or consumption paths (y_{Lsi}, c_{si}, $s < t$) followed by individual i up to the time of observation.[23]

For plausible joint distributions $G_t(w_{ti}, b_{ti}^*)$, the PPVR inheritance share φ_t will typically fall somewhere in the interval $[\varphi_t^M, \varphi_t^{KS}]$. There is, however, no theoretical reason why it should be so in general. Imagine, for instance, an economy where inheritors consume their bequests the very day they receive them, and never save afterward, so that wealth accumulation entirely comes from the savers, who never received any bequest (or negligible amounts) and who patiently accumulate savings from their labor income. Then with our definition $\varphi_t = 0\%$: in this economy, 100% of wealth accumulation comes from savings, and nothing at all comes from inheritance.

However, with the Modigliani and Kotlikoff–Summers definitions, the inheritance shares φ_t^M and φ_t^{KS} could be arbitrarily large.

[23] Of course, more data are better. If we also have (or estimate) labor income or consumption paths, then one can compute lifetime individual savings rate s_{Bti}, that is, the share of lifetime resources that was not consumed up to time t: $s_{Bti} = w_{ti} / (b_{ti}^* + y_{Lti}^*) = 1 - c_{ti}^* / (b_{ti}^* + y_{Lti}^*)$ with $y_{Lti}^* = \int_{s<t} y_{Lsi} e^{r(t-s)} ds$ = capitalized value at time t of past labor income flows, and $c_{ti}^* = \int_{s<t} c_{si} e^{r(t-s)} ds$ = capitalized value at time t of past consumption flows. By definition, inheritors are individuals who consumed more than their labor income (i.e., $w_{ti} < b_{ti}^* \leftrightarrow c_{ti}^* > y_{Lti}^*$), while savers are individuals who consumed less than their labor income (i.e., $w_{ti} \geq b_{ti}^* \leftrightarrow c_{ti}^* \leq y_{Lti}^*$). But the point is that we only need to observe an individual's wealth (w_{ti}) and capitalized inheritance (b_{ti}^*) to determine whether he or she is an inheritor or a saver, and in order to compute the share of inherited wealth.

15.4.1.5 A Simplified Definition: Inheritance Flows versus Saving Flows

When available micro data is not sufficient to apply the PPVR definition, one can also use a simplified, approximate definition based on the comparison between inheritance flows and saving flows.

Assume that all we have is macro data on inheritance flows $b_{yt} = B_t / Y_t$ and savings flows $s_t = S_t / Y_t$. Suppose for simplicity that both flows are constant over time: $b_{yt} = b_y$ and $s_t = s$. We want to estimate the share $\varphi = W_B / W$ of inherited wealth in aggregate wealth. The difficulty is that we typically do not know which part of the aggregate saving rate s comes from the return to inherited wealth, and which part comes from labor income (or from the return to past savings). Ideally, one would like to distinguish between the savings of inheritors and self-made individuals (defined along the lines explained above), but this requires micro data over two generations. In the absence of such data, a natural starting point would be to assume that the propensity to save is on average the same whatever the income sources. That is, a fraction $\varphi \cdot \alpha$ of the saving rate s should be attributed to the return to inherited wealth, and a fraction $1 - \alpha + (1 - \varphi) \cdot \alpha$ should be attributed to labor income (and to the return to past savings), where $\alpha = Y_K / Y$ is the capital share in national income and $1 - \alpha = Y_L / Y$ is the labor share. Assuming again that we are in steady state, we obtain the following simplified formula for the share of inherited wealth in aggregate wealth:

$$\varphi = \frac{b_y + \varphi \cdot \alpha \cdot s}{b_y + s},$$

$$\text{i.e.,} \varphi = \frac{b_y}{b_y + (1 - \alpha) \cdot s}.$$

Intuitively, this formula simply compares the size of the inheritance and saving flows. Because all wealth must originate from one of the two flows, it is the most natural way to estimate the share of inherited wealth in total wealth.[24]

There are a number of caveats with this simplified formula. First, real-world economies are generally out of steady state, so it is important to compute average values of b_y, s, and α over relatively long periods of time (typically over the past H years, with $H = 30$ years). If one has time-series estimates of the inheritance flow b_{ys}, capital share α_s, and saving rate s_s, then one can use the following full formula, which capitalizes past inheritance and savings flows at rate $r - g$:

[24] Similar formulas based on the comparison of inheritance and saving flows have been used by DeLong (2003) and Davies et al. (2012, pp. 123–124). One important difference is that these authors do not take into account the fact that the savings flow partly comes from the return to inherited wealth. We return to this point in Section **15.5.4**.

$$\varphi = \frac{\displaystyle\int_{t-H \le s \le t} e^{(r-g)(t-s)} \cdot b_{ys} \cdot ds}{\displaystyle\int_{t-H \le s \le t} e^{(r-g)(t-s)} \cdot \left(b_{ys} + (1-\alpha_s) \cdot s_s\right) \cdot ds}.$$

With constant flows, the full formula boils down to $\varphi = \frac{b_y}{b_y + (1-\alpha) \cdot s}$.

Second, one should bear in mind that the simplified formula $\varphi = b_y / (b_y + (1-\alpha) \cdot s)$ is an approximate formula. In general, as we show below, it tends to underestimate the true share of inheritance, as computed from micro data using the PPVR definition. The reason is that individuals who have only labor income tend to save less (in proportion to their total income) than those who have large inherited wealth and capital income, which. in turn, seems to be related to the fact that wealth (and particularly inherited wealth) is more concentrated than labor income.

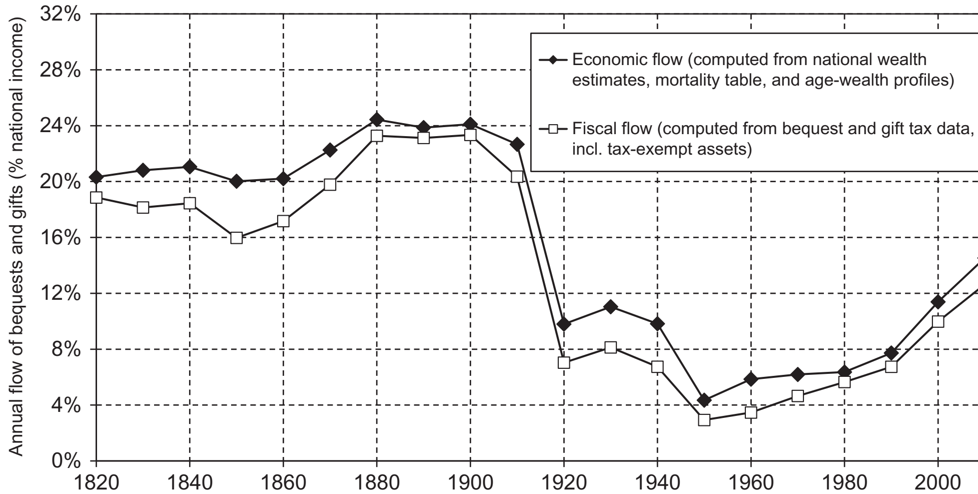
On the positive side, simplified estimates of φ seem to follow micro-based estimates relatively closely (much more closely than KSM estimates, which are either far too small or far too large), and they are much less demanding in terms of data. One only needs to estimate macro flows. Another key advantage of the simplified definition over KSM definitions is that it does not depend upon the sensitive choice of the rate of return or the rate of capital gains or losses. Whatever these rates might be, they should apply equally to inherited and self-made wealth (at least as a first approximation), so one can simply compare inheritance and saving flows.

15.4.2 The Long-Run Evolution of Inheritance in France 1820–2010

15.4.2.1 The Inheritance Flow–National Income Ratio b_{yt}

What do we empirically know about the historical evolution of inheritance? We start by presenting the evidence on the dynamics of the inheritance to national income ratio b_{yt} in France, a country for which, as we have seen in Section 15.3, historical data sources are exceptionally good (Piketty, 2011). The main conclusion is that b_{yt} has followed a spectacular U-shaped pattern over the twentieth century. The inheritance flow was relatively stable, around 20–25% of national income throughout the 1820–1910 period (with a slight upward trend), before being divided by a factor of about 5–6 between 1910 and the 1950s, and then multiplied by a factor of about 3–4 between the 1950s and the 2000s (see Figure 15.17).

These are enormous historical variations, but they appear to be well-founded empirically. In particular, the patterns for b_{yt} are similar with two independent measures of the inheritance flow. The first, what we call the fiscal flow, uses bequest and gift tax data and makes allowances for tax-exempt assets such as life insurance. The second measure, what we call the economic flow, combines estimates of private wealth W_t, mortality tables, and observed age–wealth profile, using the following accounting equation:



The annual inheritance flow was about 20−25% of national income during the nineteenth century and until 1914; it then fell to less than 5% in the 1950s and returned to about 15% in 2010

Figure 15.17 The annual inheritance flow as a fraction of national income, France 1820–2010.

$$B_t^* = (1+v_t)\cdot\mu_t\cdot m_t\cdot W_t,$$

where m_t = mortality rate (number of adult decedents divided by total adult population), μ_t = ratio between average adult wealth at death and average adult wealth for the entire population, and $v_t = V_t/B_t$ = estimate of the gift/bequest flow ratio.

The gap between the fiscal and economic flows can be interpreted as capturing tax evasion and other measurement errors. It is approximately constant over time and relatively small, so that the two series deliver consistent long-run patterns (see Figure 15.17).

The economic flow series allow—by construction—for a straightforward decomposition of the various effects at play in the evolution of b_{yt}. In the above equation, dividing both terms by Y_t we get

$$b_{yt} = B_t^*/Y_t = (1+v_t)\cdot\mu_t\cdot m_t\cdot\beta_t.$$

Similarly, dividing by W_t we can define the rate of wealth transmission b_{wt} as

$$\begin{aligned} b_{wt} &= B_t^*/W_t = (1+v_t)\cdot\mu_t\cdot m_t = \mu_t^*\cdot m_t \\ \text{with}\,\mu_t^* &= (1+v_t)\cdot\mu_t = \text{gift-corrected ratio} \end{aligned}$$

If $\mu_t = 1$ (i.e., decedents have the same average wealth as the living) and $v_t = 0$ (no gift), then the rate of wealth transmission is simply equal to the mortality rate: $b_{wt} = m_t$ (and $b_{yt} = m_t\cdot\beta_t$). If $\mu_t = 0$ (i.e., decedents die with zero wealth, such as in Modigliani's pure life-cycle theory of wealth accumulation) and $v_t = 0$ (no gift), then there is no inheritance at all: $b_{wt} = b_{yt} = 0$.

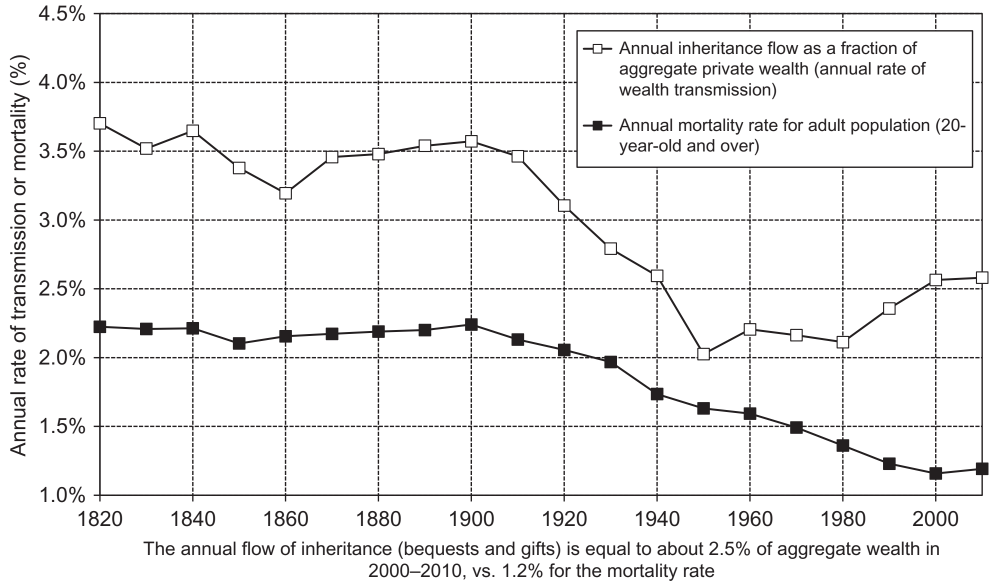


Figure 15.18 Inheritance flow versus mortality rate, France 1820–2010.

Using these accounting equations, we can see that the U-shaped pattern followed by the French inheritance–income ratio b_{yt} is the product of two U-shaped evolutions. First, it partly comes from the U-shaped evolution of the private wealth–income ratio β_t. The U-shaped evolution of b_{yt}, however, is almost twice as marked at that of β_t. The wealth–income ratio was divided by a factor of about 2–3 between 1910 and 1950 (from 600–700% to 200–300%, see Figure 15.2), whereas the inheritance flow was divided by a factor around 5–6 (from 20–25% to about 4%, see Figure 15.17). The explanation is that the rate of wealth transmission $b_{wt}=\mu_t^* \cdot m_t$ has also been following a U-shaped pattern: it was almost divided by 2 between 1910 and 1950 (from over 3.5% to just 2%), and it has been rising again to about 2.5% in 2010 (see Figure 15.18).

The U-shaped pattern followed by b_{wt}, in turn, entirely comes from μ_t^*. The relative wealth of decedents was at its lowest historical level in the aftermath of World War II (which, as we shall see below, is largely due to the fact that it was too late for older cohorts to recover from the shocks and reaccumulate wealth after the war). Given that aggregate wealth was also at its lowest historical level, the combination of these two factors explains the exceptionally low level of the inheritance flow in the 1950s–1960s. By contrast, the mortality rate m_t has been constantly diminishing: this long-run downward trend is the mechanical consequence of the rise in life expectancy (for a given cohort size).[25]

[25] The mortality rate, however, is about to rise somewhat in coming decades in France owing to the aging of the baby boomers (see Piketty, 2011). This effect will be even stronger in countries where cohort size has declined in recent decades (such as Germany or Japan) and will tend to push inheritance flows toward even higher levels.

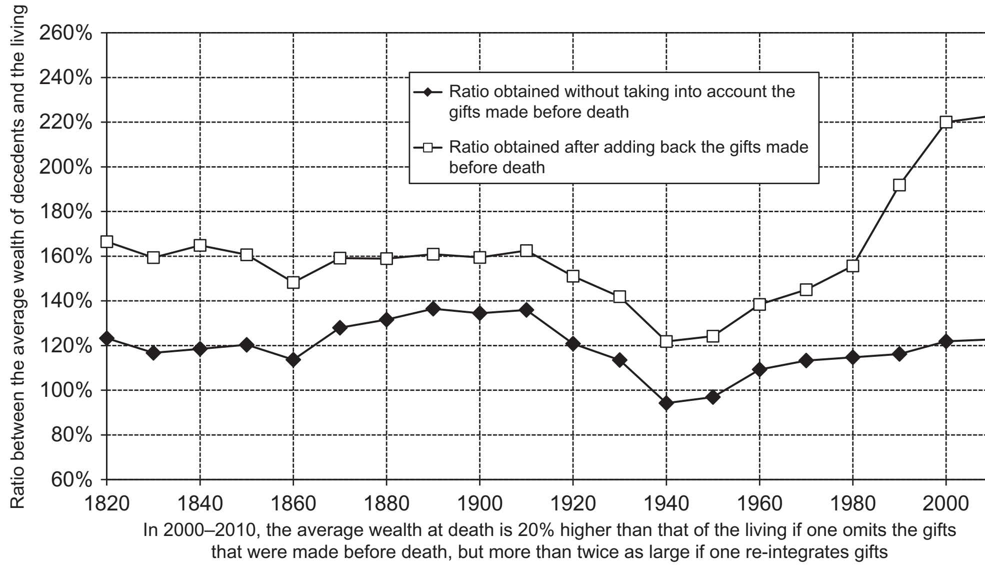


Figure 15.19 The ratio between average wealth at death and average wealth of the living, France 1820–2010.

In the recent decades, a very large part of the rise in $\mu_t^* = (1 + \upsilon_t) \cdot \mu_t$ comes from the rise in the gift–bequest ratio υ_t, which used to be about 20% during most of the nineteenth to twentieth centuries, and has gradually risen to as much as 80% in recent decades (see Figure 15.19). That is, the gift flow is currently almost as large as the bequest flow.

Although there is still much uncertainty about the reasons behind the rise in gifts, the evidence suggests that it started before the introduction of new tax incentives for gifts in the 1990s–2000s, and has more to do with the growing awareness by wealthy parents that they will die old and that they ought to transmit part of their wealth inter-vivos if they want their children to fully benefit from it.

In any case, one should not underestimate the importance of gifts. In particular, one should not infer from a declining age–wealth profile at old ages or a relatively low relative wealth of decedents that inheritance is unimportant: this could simply reflect the fact that decedents have already given away a large part of their wealth.

15.4.2.2 The Inheritance Stock-Aggregate Wealth Ratio φ_t

How do the annual inheritance flows transmit into cumulated inheritance stocks? Given the data limitations we face, we show on Figure 15.20 two alternative estimates for the share φ_t of total inherited wealth in aggregate French wealth between 1850 and 2010. According to both measures, there is again a clear U-shaped pattern. The share of inherited wealth φ_t was as large as 80–90% of aggregate wealth in 1850–1910, down to as little as 35–45% around 1970, and back up to 65–75% by 2010.

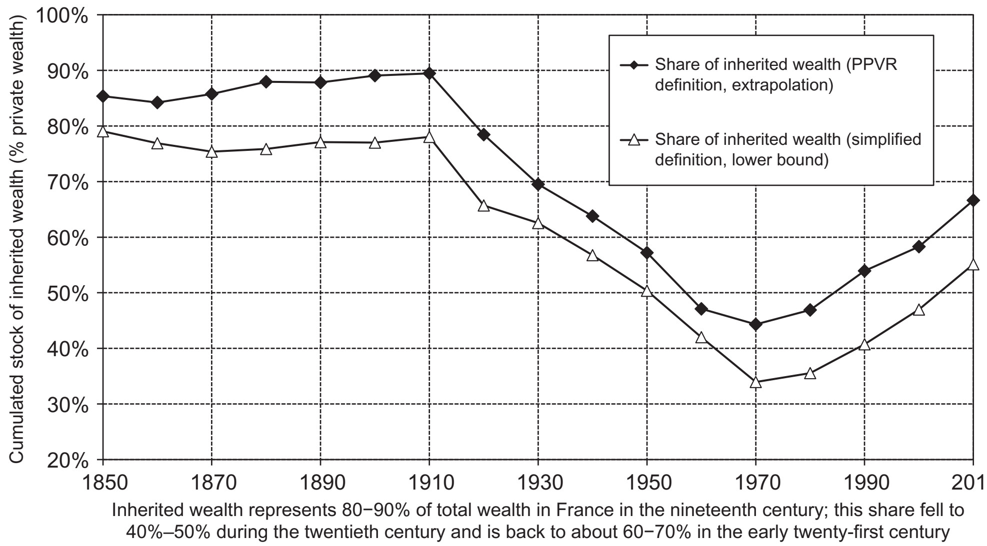


Figure 15.20 The cumulated stock of inherited wealth as a fraction of aggregate private wealth, France 1850–2010.

The higher series, which we see as the most reliable, was obtained by applying the micro-based PPVR definition (see Section 15.4.1.4). The limitation here is that the set of micro data on wealth over two generations that has been collected in French historical archives is more complete for Paris than for the rest of France (see Piketty et al., 2006, 2013). For years with missing data for the rest of France, the estimates reported on Figure 15.20 were extrapolated on the basis of the Parisian data. Ongoing data collection suggests that the final estimates will not be too different from the approximate estimates reported here.

The lower series, which we see as a lower bound, comes from the simplified definition based on the comparison of inheritance and saving flows (see Section 15.4.1.5).[26] The key advantage of this simplified definition is that it requires much less data: it can readily be computed from the inheritance flow series b_{yt} that were reported above. It delivers estimates of the inheritance share φ_t that are always somewhat below the micro-based estimates, with a gap that appears to be approximately constant. The gap seems to be due to the fact that the simplified definition attributes too much saving to pure labor earners with little inheritance.

In both series, the share φ_t of total inherited wealth in aggregate wealth reaches its lowest historical point in the 1970s, whereas the inheritance flow b_{yt} reaches its lowest point in the immediate aftermath of World War II. The reason is that the stock of

[26] The series was computed as $\varphi = b_y/(b_y + (1-\alpha)\cdot s)$ using 30-year averages for saving rates, capital shares, and inheritance flows.

inherited wealth comes from cumulating the inheritance flows of the previous decades—hence the time lag.

15.4.3 Evidence from Other Countries

What do we know about the importance of inheritance in countries other than France? A recent wave of research attempts to construct estimates of the inheritance flow–national income ratio b_{yt} in a number of European countries. The series constructed by Atkinson (2013) for Britain and Schinke (2013) for Germany show that b_{yt} has also followed a U-shaped pattern in these two countries over the past century (see Figure 15.21). Data limitations, however, make it difficult at this stage to make precise comparisons between countries.

For Britain, the inheritance flow b_{yt} of the late nineteenth to early twentieth centuries seems to be similar to that of France, namely about 20–25% of national income. The flow then falls following the 1914–1945 shocks, albeit less spectacularly than in France, and recovers in recent decades. Karagiannaki (2011), in a study of inheritance in the United Kingdom from 1984 to 2005, also finds a marked increase in that period. The rebound, however, seems to be less strong in Britain than in France, so that the inheritance flow appears smaller than in France today. We do not know yet whether this finding is robust. Available British series are pure "fiscal flow" series (as opposed to French series, for which we have both an "economic" and a "fiscal" estimate). As pointed out by Atkinson (2013), the main reason for the weaker British rebound in recent decades is that the gift–bequest ratio v_t has not increased at all according to fiscal data (v_t has remained relatively flat at a

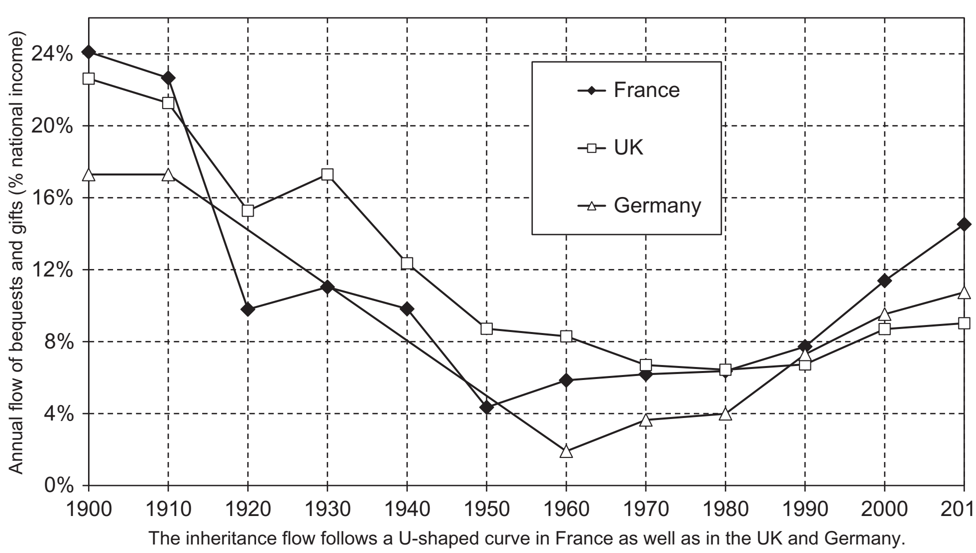


Figure 15.21 The inheritance flow in Europe, 1900–2010.

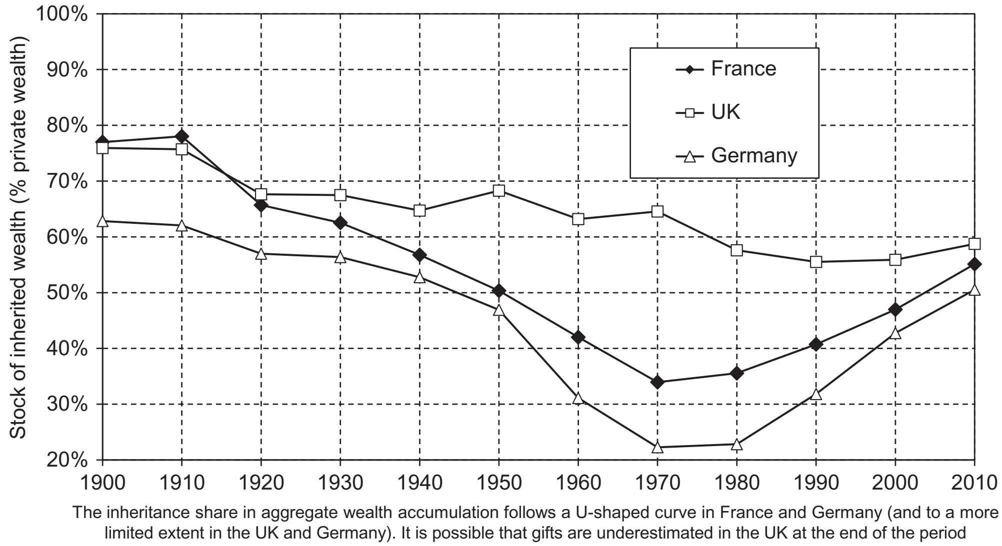


Figure 15.22 The inheritance stock in Europe, 1900–2010.

low level, around 10–20%). According to Atkinson, this could be due to substantial underreporting of gifts to tax authorities.

Germany also exhibits a U-shaped pattern of inheritance flow b_{yt} that seems to be broadly as sharp as in France. In particular, just as in France, the strong German rebound in recent decades comes with a large rise in the gift–bequest ratio ν_t during the 1990s–2000s (υ_t is above 50–60% in the 2000s). The overall levels of b_{yt} are generally lower in Germany than in France, which given the lower aggregate wealth–income ratio β_t is not surprising. Should we compare the rates of wealth transmission (i.e., $b_{wt} = b_{yt}/\beta_t$), then the levels would be roughly the same in both countries in 2000–2010.

We report on Figure 15.22 the corresponding estimates for the share φ_t of total inherited wealth in aggregate wealth, using the simplified definition $\varphi = b_y/(b_y + (1-\alpha)s)$. For Germany, the inheritance share φ_t appears to be generally smaller than in France. In particular, it reaches very low levels in the 1960s–1970s, owing to the extremely low inheritance flows in Germany in the immediate postwar period, and to large saving rates. In recent decades, the German φ_t has been rising fast and seems to catch up with France's. In the United Kingdom, the inheritance share φ_t apparently never fell to the low levels observed in France and Germany in the 1950s, and seems to be always higher than on the Continent. The reason, for the recent period, is that the United Kingdom has had relatively low saving rates since the 1970s.[27]

[27] In effect, British saving rates in recent decades are insufficient to explain the large rise in the aggregate wealth–income ratio, which can only be accounted for by large capital gain (Piketty and Zucman, 2014). The simplified definition of φ_t based on the comparison between inheritance and saving flows assumes the same capital gains for inherited and self-made wealth.

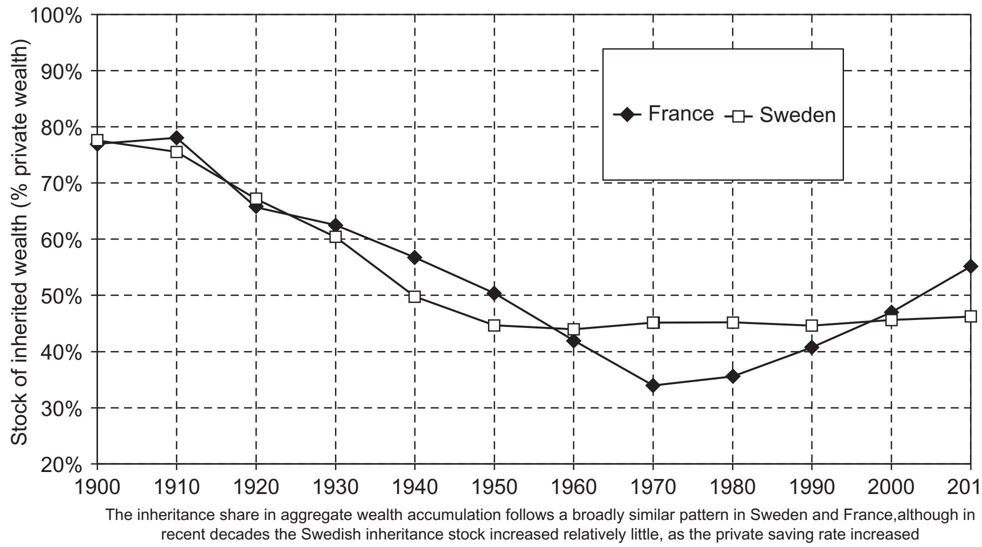


Figure 15.23 The inheritance stock in France and Sweden, 1900–2010.

Recent historical research suggests that inheritance flows have also followed U-shaped patterns in Sweden (see Ohlsson et al., 2013). Here b_{yt} appears to be smaller than in France, but this again seems largely due to lower β_t ratios. When we look at the implied b_{wt} and φ_t ratios, which in a way are the most meaningful ratios to study, then both the levels and shape are relatively similar across European countries. As shown by Figure 15.23, the share of inherited wealth followed the same evolution in Sweden and France in the twentieth century (the main difference is that it seems to have increased a bit less in Sweden than in France in recent decades, because of a rise in the private saving rate). We stress again, however, that a lot more data needs to be collected—and to some extent is currently being collected—on the historical evolution of inheritance before we can make proper international comparisons.

Prior to the recent inheritance flow estimates surveyed above, a first wave of research, surveyed by Davies and Shorrocks (1999), mostly focused on the United States, with conflicting results—the famous Modigliani–Kotlikoff–Summers controversy. More recently, Kopczuk and Edlund (2009) observe that in estate tax data, the share of women among the very wealthy in the United States peaked in the late 1960s (at nearly one-half) and then declined to about one-third. They argue that this pattern reflects changes in the importance of inheritance, as women are less likely to be entrepreneurs. Wolff and Gittleman (2013) analyze Survey of Consumer Finances (SCF) data and find little evidence of a rise in inheritances since the late 1980s. Looking at *Forbes*' data, Kaplan and Rauh (2013) find that Americans in the Forbes 400 are less likely to have inherited their wealth today than in the 1980s. It is unclear, however, whether this result reflects a true economic phenomenon or illustrates the limits of *Forbes* and other wealth rankings.

Inherited wealth holdings are probably harder to spot than self-made wealth, first because inheritors' portfolios tend to be more diversified, and also because inheritors may not like to be in the press, while entrepreneurs usually enjoy it and do not attempt to dissimulate their wealth nearly as much. The conclusions about the relative importance of inherited versus self-made wealth obtained by analyzing *Forbe*s list data may thus be relatively fragile.

In the end, there remain important uncertainties about the historical evolution of inheritance in the United States. There are reasons to believe that inheritance has historically been less important in the United States than in Europe, because population growth has been much larger (more on this below). It is unclear whether this still applies today, however. Given the relatively low U.S. saving rates in recent decades, it is possible that even moderate inheritance flows imply a relatively large share φ_t of total inherited wealth in aggregate wealth (at least according to the simplified definition of φ based on the comparison between b_y and s).

One difficulty is that U.S. fiscal data on bequests and gifts are relatively low quality (in particular because the federal estate tax only covers few decedents; in 2012 only about 1 decedent out of 1000 pays the estate tax). One can use survey data (e.g., from the SCF) to estimate the relative wealth of decedents μ_t and compute the economic inheritance flow $b_{yt}=(1+\upsilon_t)\cdot\mu_t\cdot\beta_t$. One key problem is that one needs to find ways to estimate the gift–bequest ratio υ_t, which is not easy to do in the absence of high-quality fiscal data. Self-reported retrospective data on bequest and gift receipts usually suffer from large downward biases and should be treated with caution. In countries where there exists exhaustive administrative data on bequests and gifts (such as France, and to some extent Germany), survey-based self-reported flows appear to be less than 50% of fiscal flows. This may contribute to explain the low level of inheritance receipts found by Wolff and Gittleman (2013).[28]

15.5. ACCOUNTING FOR THE EVIDENCE: MODELS AND PREDICTIONS

15.5.1 Shocks Versus Steady States

How can we account for the historical evidence on the evolution of the aggregate wealth–income ratio, the concentration of wealth, and the share of inherited wealth? In this section, we describe the theoretical models that have been developed to address this question. While we still lack a comprehensive model able to rigorously and

[28] One additional challenge in this study is that inherited assets are generally valued using asset prices at the time the assets were transmitted: no capital gain is included—which probably contributes to a relatively low estimated inheritance share in total U.S. wealth (about 20%, just like in Modigiani's estimates). A comparison between inheritance flows and saving flows (using the simplified formula) would likely lead to more balanced results.

quantitatively assess the various effects at play, the literature makes it possible to highlight some of the key forces.

We are primarily concerned here about the determinants of long-run steady states. In practice, as should be clear from the historical series presented above, real-world economies often face major shocks and changes in fundamental parameters, so that we observe large deviations from steady states. In particular, the large decline in the aggregate wealth–income ratios β_t between 1910 and 1950 is due to the shocks induced by the two World Wars. By using detailed series on saving flows and war destructions, one can estimate the relative importance of the various factors at play (Piketty and Zucman, 2014). In the case of France and Germany, three factors of comparable magnitude each account for approximately one-third of the total 1910–1950 fall of β_t: insufficient national savings (a large part of private saving was absorbed by public deficits); war destructions; and the fall of relative assets prices (real estate and equity prices were historically low in 1950–1960, partly due to policies such as rent control and nationalization). In the case of Britain, war destructions were relatively minor, and the other two factors each account for about half of the fall in the ratio of wealth to income (war-induced public deficits were particularly large).[29]

In thinking about the future, is the concept of a steady state a relevant point of reference? Historical evidence suggests that it is. Whereas the dynamics of wealth and inequality has been chaotic in the twentieth century, eighteenth- and nineteenth-century United Kingdom and France can certainly be analyzed as being in a steady state characterized by low-growth, high wealth–income ratios, high levels of wealth concentration, and inheritance flows. This is true despite the fact that there were huge changes in the nature of wealth and of economic activity (from agriculture to industry).[30] The shocks of the twentieth century put an end to this steady state, and it seems justified to ask: if countries are to converge to a new steady state in the twenty-first century (that is, if the shocks of the twentieth century do not happen again), which long-term ratios will they reach?

We show that over a wide range of models, the long-run magnitude and concentration of wealth and inheritance are a decreasing function of g and an increasing function of $\bar{r}$, where g is the economy's growth rate and $\bar{r}$ is the net-of-tax rate of return to wealth. That is, under plausible assumptions, both the wealth–income ratio and the concentration of wealth tend to take higher steady-state values when the long-run growth rate is lower and when the net-of-tax rate of return is higher. In particular, a higher $\bar{r} - g$ tends to magnify steady-state wealth inequalities. Although there does not exist yet any

[29] For detailed decompositions of private and national wealth accumulation over the various subperiods, see Piketty and Zucman (2014).

[30] In particular, private wealth/income ratios and inheritance flows seemed quite stable in nineteenth-century France (with perhaps a slight upward trend at the end of the century), despite major structural economic changes. This suggests that although the importance of inheritance and wealth may rise and fall in response to the waves of innovation, a steady-state analysis is a fruitful perspective.

rigorous calibrations of these theoretical models, we argue that these predictions are broadly consistent with both the time-series and cross-country evidence. These findings also suggest that the current trends toward rising wealth–income ratios and wealth inequality might continue during the twenty-first century, both because of population and productivity-growth slowdown, and because of rising international competition to attract capital.

15.5.2 The Steady-State Wealth–Income Ratio: $\beta = s/g$

The most useful steady-state formula to analyze the long-run evolution of wealth–income and capital–output ratios is the Harrod–Domar–Solow steady-state formula:

$$\beta_t \rightarrow \beta = s/g.$$

With s = long-run (net-of-depreciation) saving rate, g = long-run growth rate.[31]

The steady-state formula $\beta = s/g$ is a pure accounting equation. By definition, it holds in the steady state of any micro-founded, one-good model of capital accumulation, independently of the exact nature of saving motives. It simply comes from the wealth-accumulation equation $W_{t+1} = W_t + S_t$, which can be rewritten in terms of wealth–income ratio $\beta_t = W_t/Y_t$:

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

With $1 + g_t = Y_{t+1}/Y_t$ = growth rate of national income, $s_t = S_t/Y_t$ = net saving rate.

It follows immediately that if $s_t \rightarrow s$ and $g_t \rightarrow g$, then $\beta_t \rightarrow \beta = s/g$.

The Harrod–Domar–Solow says something trivial but important in a low-growth economy, the sum of capital accumulated in the past can become very large, as long as the saving rate remains sizable.

For instance, if the long-run saving rate is $s = 10\%$, and if the economy permanently grows at rate $g = 2\%$, then in the long run the wealth–income ratio has to be equal to $\beta = 500\%$, because it is the only ratio such that wealth rises at the same rate as income: $s/\beta = 2\% = g$. If the long-run growth rate declines to $g = 1\%$, and the economy keeps saving at rate $s = 10\%$, then the long-run wealth–income ratio will be equal to $\beta = 1000\%$.

In the long run, output growth g is the sum of productivity and population growth. In the standard one-good growth model, output is given by $Y_t = F(K_t, L_t)$, where K_t is non-human capital input and L_t is human labor input (i.e., efficient labor supply). L_t can be written as the product of raw-labor supply N_t and labor productivity parameter h_t. That is, $L_t = N_t \cdot h_t$, with $N_t = N_0 \cdot (1 + n)^t$ (n is the population growth rate) and $h_t = h_0 \cdot (1 + h)^t$ (h is

[31] When one uses gross-of-depreciation saving rates rather than net rates, the steady-state formula writes $\beta = s/(g + \delta)$ with s the gross saving rate, and δ the depreciation rate expressed as a proportion of the wealth stock.

the productivity growth rate). The economy's long-run growth rate g is given by the growth rate of L_t. Therefore it is equal to $1+g=(1+n)\cdot(1+h)$, i.e., $g \approx n+h$.[32] The long-run g depends both on demographic parameters (in particular, fertility rates) and on productivity-enhancing activities (in particular, the pace of innovation).

The long-run saving rate s also depends on many forces: s captures the strength of the various psychological and economic motives for saving and wealth accumulation (dynastic, life cycle, precautionary, prestige, taste for bequests, etc.). The motives and tastes for saving vary a lot across individuals and potentially across countries. Whether savings come primarily from a life cycle or a bequest motive, the $\beta=s/g$ formula will hold in steady state. In case saving is exogenous (as in the Solow model), the long-run wealth–income ratio will obviously be a decreasing function of the income growth rate g. This conclusion, however, is also true in a broad class of micro-founded, general equilibrium models of capital accumulation in which s can be endogenous and can depend on g. That is the case, in particular, in the infinite-horizon, dynastic model (in which s is determined by the rate of time preference and the concavity of the utility function), in "bequest-in-the-utility-function" models (in which the long-run saving rate s is determined by the strength of the bequest or wealth taste), and in most endogenous growth models (see box below). In all cases, for given preference parameters, the long-run $\beta=s/g$ tends to be higher when the growth rate is lower. A growth slowdown—coming from a decrease in population or productivity growth—tends to lead to higher capital–output and wealth–income ratios.

Box: The steady-state wealth–income ratio in macro models

Dynastic Model

Assume that output is given by $Y_t=F(K_t, L_t)$, where K_t is the capital stock and L_t is efficient labor and grows exogenously at rate g. Output is either consumed or added to the capital stock. We assume a closed economy, so the wealth–income ratio is the same as the capital–output ratio. In the infinite-horizon, dynastic model, each dynasty maximizes

$$V=\int_{t\geq s} \mathrm{e}^{-\theta t} U(c_t)$$

where θ is the rate of time preference and $U(c_t)=c^{1-\gamma}(1-\gamma)$ is a standard utility function with a constant intertemporal elasticity of substitution equal to $1/\gamma$. This elasticity of substitution is often found to be small, typically between 0.2 and 0.5, and is in any case smaller than one. Therefore γ is typically bigger than one.

Continued

[32] To obtain the exact equality $g=n+h$, one needs to use instantaneous (continuous-time) growth rates rather than annual (discrete-time) growth rates. That is, with $N_t=N_0\cdot \mathrm{e}^{nt}$ (with n = population growth rate) and $h_t=h_0\cdot \mathrm{e}^{ht}$, we have $L_t=N_t\cdot h_t=L_0\cdot \mathrm{e}^{gt}$, with $g=n+h$.

The first-order condition describing the optimal consumption path of each dynasty is: $dc_t/dt = (r - \theta) \cdot c_t/\gamma$, i.e., utility-maximizing agents want their consumption path to grow at rate $g_c = (r - \theta)/\gamma$. This is a steady state if and only if $g_c = g$, i.e., $r = \theta + \gamma g$, what is known as the modified Golden Rule of capital accumulation. The long-run rate of return $r = \theta + \gamma g$ is entirely determined by preference parameters and the growth rate and is larger than g.

The steady-state saving rate is equal to $s = \alpha \cdot g/r = \alpha \cdot g/(\theta + \gamma g)$, where $\alpha = r \cdot \beta$ is the capital share. Intuitively, a fraction g/r of capital income is saved in the long run, so that dynastic wealth grows at the same rate g as national income. The saving rate $s = s(g)$ is an increasing function of the growth rate, but rises less fast than g, so that the steady-state wealth–income ratio $\beta = s/g$ is a decreasing function of the growth rate.

For instance, with a Cobb–Douglas production function (in which case the capital share is entirely set by technology and is constantly equal to α), the wealth–income ratio is given by $\beta = \alpha/r = \alpha/(\theta + \gamma \cdot g)$ and takes its maximum value $\overline{\beta} = \alpha/\theta$ for $g = 0$.

One unrealistic feature of the dynastic model is that it assumes an infinite long-run elasticity of capital supply with respect to the net-of-tax rate of return, which mechanically entails extreme consequences for optimal capital tax policy (namely, zero tax). The "bequest-in-the-utility-function" model provides a less extreme and more flexible conceptual framework in order to analyze the wealth accumulation process.

Wealth-in-the-Utility-Function Model

Consider a dynamic economy with a discrete set of generations 0, 1, ..., t, ..., zero population growth, and exogenous labor productivity growth at rate $g > 0$. Each generation has measure $N_t = N$, lives one period, and is replaced by the next generation. Each individual living in generation t receives bequest $b_t = w_t \geq 0$ from generation $t-1$ at the beginning of period t, inelastically supplies one unit of labor during his lifetime (so that labor supply $L_t = N_t = N$), and earns labor income y_{Lt}. At the end of period, he then splits lifetime resources (the sum of labor income and capitalized bequests received) into consumption c_t and bequests left $b_{t+1} = w_{t+1} \geq 0$, according to the following budget constraint:

$$c_t + b_{t+1} \leq y_t = y_{Lt} + (1 + r_t)b_t$$

The simplest case is when the utility function is defined directly over consumption c_t and the increase in wealth $\Delta w_t = w_{t+1} - w_t$ and takes a simple Cobb–Douglas form: $V(c, \Delta w) = c^{1-s}\Delta w^s$. (Intuitively, this corresponds to a form of "moral" preferences where individuals feel that they cannot possibly leave less wealth to their children than what they have received from their parents, and derive utility from the increase in wealth, maybe because this is a signal of their ability or virtue.) Utility maximization then leads to a fixed saving rate: $w_{t+1} = w_t + s y_t$. By multiplying per capita values by population $N_t = N$ we have the same linear transition equation at the aggregate level: $W_{t+1} = W_t + sY_t$. The long-run wealth–income ratio is given by $\beta_t \to \beta = s/g$. It depends on the strength of the bequest motive and on the rate of productivity growth.

With other functional forms for the utility function, e.g., with $V = V(c, w)$, or with heterogenous labor productivities or saving tastes across individuals, one simply needs to replace the parameter s by the properly defined average wealth or bequest taste parameter. For instance, with $V(c, w) = c^{1-s}w^s$, utility maximization leads to $w_{t+1} = s \cdot (w_t + y_t)$ and

Continued

$\beta_t \to \beta = s/(g+1-s) = \widetilde{s}/g$, with $\widetilde{s} = s(1+\beta) - \beta$ the conventional saving rate (i.e., defined relative to income). See Section 15.5.4.1 for a simple application of this model to the analysis of the steady-state distribution of wealth.

Endogenous Growth Models

In endogenous growth models with imperfect international capital flows, the growth rate might rise with the saving rate, but it will usually rise less than proportionally. It is only in what is known as the AK closed-economy model that the growth rate rises proportionally with the saving rate. To see this, assume zero population growth ($n=0$) and a Cobb–Douglas production function $Y = K^{\alpha} \cdot (A_L \cdot L)^{1-\alpha}$. Further assume that the productivity parameter is endogenously determined by an economy-wide capital accumulation externality, such that $A_L = A_0 \cdot K$. Then we have $Y = A \cdot K$, with $A = (A_0 \cdot L_0)^{1-\alpha}$. For a given saving rate $s > 0$, the growth rate is given by $g = g(s) = s \cdot A$. The growth rate rises proportionally with the saving rate, so that the wealth–income ratio is entirely set by technology: $\beta = s/g = 1/A$ is a constant.

In more general endogenous growth models, the rate of productivity growth depends not only on the pace of capital accumulation, but also—and probably more importantly—on the intensity of innovation activities, the importance of education spendings, the position on the international technological frontier, and a myriad of other policies and institutions, so that the resulting growth rate rises less than proportionally with the saving rate.

The slowdown of income growth is the central force explaining the rise of wealth–income ratios in rich countries over the 1970–2010 period, particularly in Europe and Japan, where population growth has slowed markedly (and where saving rates are still high relative to the United States). As Piketty and Zucman (2014) show, the cumulation of saving flows explains the 1970–2010 evolution of β in the main rich countries relatively well. An additional explanatory factor over this time period is the gradual recovery of relative asset prices. In the very long run, however, relative asset-price movements tend to compensate each other, and the one-good capital accumulation model seems to do a good job at explaining the evolution of wealth–income ratios.

It is worth stressing that the $\beta = s/g$ formula works both in closed-economy and open-economy settings. The only difference is that wealth–income and capital–output ratios are the same in closed-economy settings but can differ in open-economy environments.

In the closed-economy case, private wealth is equal to domestic capital: $W_t = K_t$.[33] National income Y_t is equal to domestic output $Y_{dt} = F(K_t, L_t)$. Saving is equal to domestic investment, and the private wealth–national income ratio $\beta_t = W_t/Y_t$ is the same as the domestic capital–output ratio $\beta_{kt} = K_t/Y_{dt}$.

In the open economy case, countries with higher saving rates $s_a > s_b$ accumulate higher wealth ratios $\beta_a = s_a/g > \beta_b = s_b/g$ and invest some their wealth in countries with

[33] For simplicity we assume away government wealth and saving.

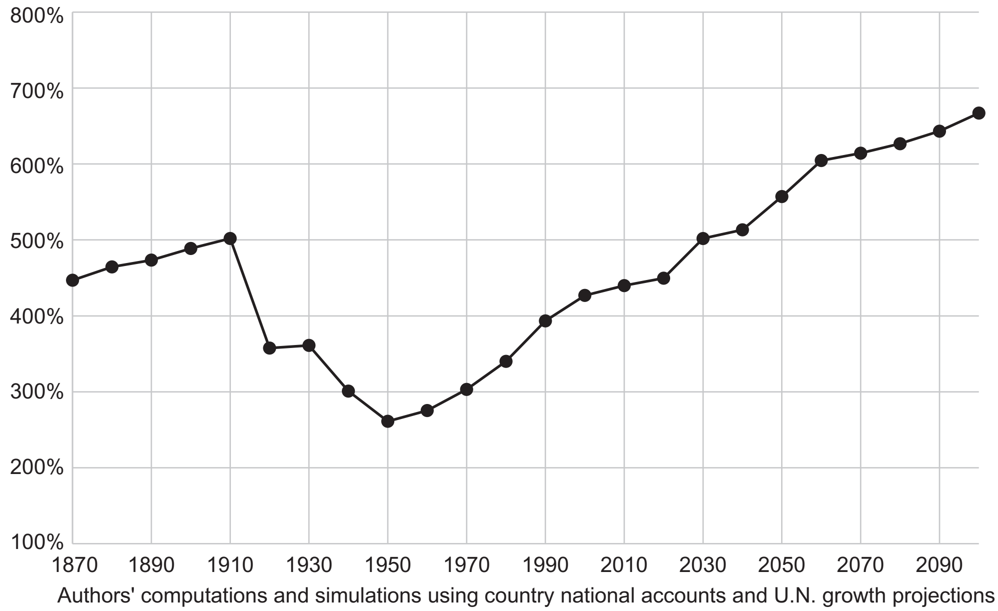


Figure 15.24 World wealth/national income ratio, 1870–2100.

lower saving rates, so that the capital–output ratio is the same everywhere (assuming perfect capital mobility). Noting N_a and N_b the population of countries a and b, $N = N_a + N_b$ the world population, $Y = Y_a + Y_b$ the world output, and $s = (s_a \cdot Y_a + s_b \cdot Y_b)/Y$ the world saving rate, and assuming that each country's effective labor supply is proportional to population and grows at rate g, then the long-run wealth–income and capital–output ratio at the world level will be equal to $\beta = s/g$. With perfect capital mobility, each country will operate with the same capital–output ratio $\beta = s/g$. Country a with wealth $\beta_a > \beta$ will invest its extra wealth $\beta_a - \beta$ in country b with wealth $\beta_b < \beta$. Both countries have the same per capita output $y = Y/N$, but country a has a permanently higher per capita national income $y_a = y + r \cdot (\beta_a - \beta) > y$, while country b has a permanently lower per capita national income $y_b = y - r \cdot (\beta - \beta_b) < y$. In the case of Britain and France at the eve of World War I, the net foreign wealth position $\beta_a - \beta$ was of the order of 100–200%, the return on net foreign assets was about $r = 5\%$, so that national income was about 5–10% larger than domestic output.

At the world level, wealth–income and capital–output ratios always coincide (by definition). The long-run ratio is governed by the steady-state condition $\beta = s/g$. In the very long run, if the growth rate slows down at the global level (in particular due to the possible stabilization of world population), then the global β might rise. We report on Figure 15.24 one possible evolution of the world wealth–income ratio in the twenty-first century, assuming that the world-income growth rate stabilizes at about 1.5% and world saving rate at about 12%. Under these (arguably specific and uncertain) assumptions, the world β would rise to about 700–800% by the end of the twenty-first century.

15.5.3 The Steady-State Capital Share: $\alpha = r \cdot \beta = a \cdot \beta^{\frac{\sigma-1}{\sigma}}$

How does the evolution of the capital–income ratio β relate to the evolution of the capital share $\alpha_t = r_t \cdot \beta_t$ (where r_t is the average rate of return)? All depends on whether the capital–labor elasticity of substitution σ is larger or smaller than one.

Take a CES production function $Y = F(K, L) = \left(a \cdot K^{\frac{\sigma-1}{\sigma}} + (1-a) \cdot L^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma-1}{\sigma}}$. The rate of return is given by $r = F_K = a\beta^{-1/\sigma}$ (with $\beta = K/Y$), and the capital share is given by $\alpha = r \cdot \beta = a \cdot \beta^{\frac{\sigma-1}{\sigma}}$. If $\sigma > 1$, then as β_t rises, the fall of the marginal product of capital r_t is smaller than the rise of β_t, so that the capital share $\alpha_t = r_t \cdot \beta_t$ is an increasing function of β_t. Conversely, if $\sigma < 1$, the fall of r_t is bigger than the rise of β_t, so that the capital share is a decreasing function of β_t.[34]

As $\sigma \to \infty$, the production function becomes linear, that is, the return to capital is independent of the quantity of capital: this is like a robot economy where capital can produce output on its own. Conversely, as $\sigma \to 0$, the production function becomes putty clay, that is, the return to capital falls to zero if the quantity of capital is slightly above the fixed-proportion technology.

A special case is when the capital–labor elasticity of substitution σ is exactly equal to one: changes in r and in β exactly compensate each other so that the capital share is constant. This is the Cobb–Douglas case $F(K, L) = K^{\alpha} L^{1-\alpha}$. The capital share is entirely set by technology: $\alpha_t = r_t \cdot \beta_t = \alpha$. A higher capital–output ratio β_t is exactly compensated by a lower capital return $r_t = \alpha/\beta_t$, so that the product of the two is constant.

There is a large literature trying to estimate the elasticity of substitution between labor and capital, reviewed in Antras (2004) and Chirinko (2008); see also Karabarbounis and Neiman (2014). The range of estimates is wide. Historical evidence suggests that the elasticity of substitution σ may have risen over the development process. In the eighteenth to nineteenth centuries, it is likely that σ was less than one, particularly in the agricultural sector. An elasticity less than one would explain why countries with large quantities of land (e.g., the United States) had lower aggregate land values than countries with little land (the Old World). Indeed, when $\sigma < 1$, price effects dominate volume effects: when land is very abundant, the price of land is extremely low, and the product of the two is small. An elasticity less than 1 is exactly what one would accept in an economy in which

[34] Because we include all forms of capital assets into our aggregate capital concept K, the aggregate elasticity of substitution σ should be interpreted as resulting from both supply forces (producers shift between technologies with different capital intensities) and demand forces (consumers shift between goods and services with different capital intensities, including housing services versus other goods and services).

capital takes essentially one form only (land), as in the eighteenth and early nineteenth centuries. When there is too much of the single capital good, it becomes almost useless.

Conversely, in the twentieth century, capital shares α have tended to move in the same direction as capital–income ratios β. This fact suggests that the elasticity of substitution σ has been larger than one. Since the mid-1970s, in particular, we do observe a significant rise of capital shares α_t in rich countries (Figure 15.25). Admittedly, the rise in capital shares α_t was less marked than the rise of capital–income ratios β_t—in other words, the average return to wealth $r_t = \alpha_t/\beta_t$ has declined (Figure 15.26). But this decline is

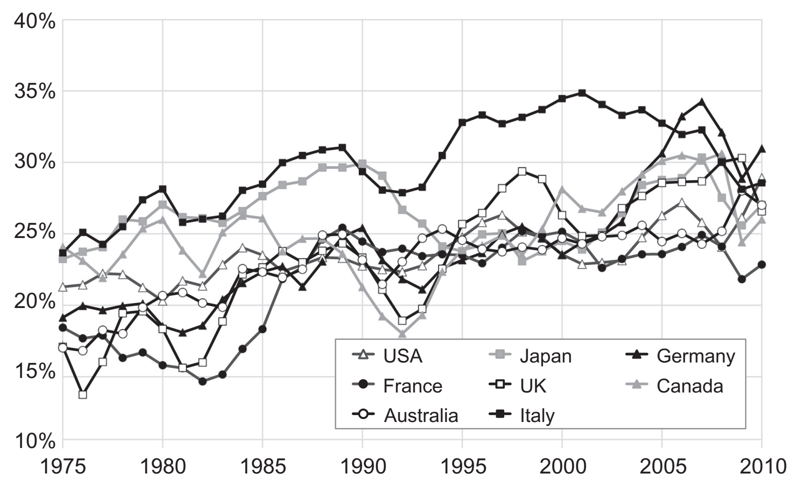


Figure 15.25 Capital shares in factor-price national income, 1975–2010.

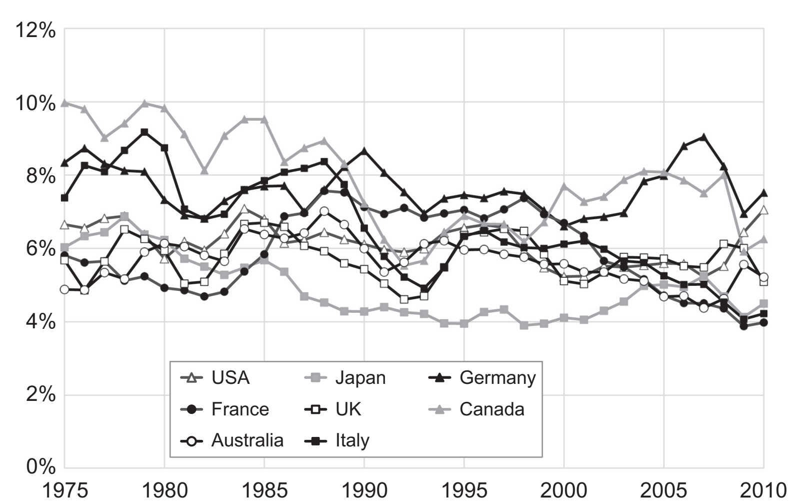


Figure 15.26 Average return on private wealth, 1975–2010.

exactly what one should expect in any economic model: when there is more capital, the rate of return to capital must go down. The interesting question is whether the average return r_t declines less or more than β_t increases. The data gathered by Piketty and Zucman (2014) suggest that r_t has declined less, i.e., that the capital share has increased, consistent with an elasticity $\sigma > 1$. This result is intuitive: an elasticity larger than one is what one would expect in a sophisticated economy with different uses for capital (not only land, but also robots, housing, intangible capital, etc.). The elasticity might even increase with globalization, as it becomes easier to move different forms of capital across borders.

Importantly, the elasticity does not need to be hugely superior to one in order to account for the observed trends. With an elasticity σ around 1.2–1.6, a doubling of capital–output ratio β can lead to a large rise in the capital share α. With large changes in β, one can obtain substantial movements in the capital share with a production function that is only moderately more flexible than the standard Cobb–Douglas function. For instance, with $\sigma = 1.5$, the capital share rises from $\alpha = 28\%$ to $\alpha = 36\%$ if the wealth–income ratio jumps from $\beta = 2.5$ to $\beta = 5$, which is roughly what has happened in rich countries since the 1970s. The capital share would reach $\alpha = 42\%$ in case further capital accumulation takes place and the wealth–income ratio attains $\beta = 8$. In case the production function becomes even more flexible over time (say, $\sigma = 1.8$), the capital share would then be as large as $\alpha = 53\%$.[35] The bottom line is that we certainly do not need to go all the way toward a robot economy ($\sigma = \infty$) in order to generate very large movements in the capital share.

15.5.4 The Steady-State Level of Wealth Concentration: *Ineq* = *Ineq* ($\bar{r} - g$)

The possibility that the capital–income ratio β—and maybe the capital share α—might rise to high levels entails very different welfare consequences depending on who owns capital. As we have seen in Section 15.3, wealth is always significantly more concentrated than income, but wealth has also become less concentrated since the nineteenth to early twentieth century, at least in Europe. The top 10% wealth holders used to own about 90% of aggregate wealth in Europe prior to World War I, whereas they currently own about 60–70% of aggregate wealth.

What model do we have to analyze the steady-state level of wealth concentration? There is a large literature devoted to this question. Early references include Champernowne (1953), Vaughan (1979), and Laitner (1979). Stiglitz (1969) is the first attempt to analyze the steady-state distribution of wealth in the neoclassical growth model. In his and similar models of wealth accumulation, there is at the same time both convergence of the macro-variables to their steady-state values and of the distribution of wealth to its steady-state form. Dynamic wealth-accumulation models with random

[35] With $a = 0.21$ and $\sigma = 1.5$, $\alpha = a \cdot \beta^{\frac{\sigma-1}{\sigma}}$ goes from 28% to 36% and 42% as β rises from 2.5 to 5 and 8. With $\sigma = 1.8$, α rises to 53% if $\beta = 8$.

idiosyncratic shocks have the additional property that a higher $\bar{r}-g$ differential (where $\bar{r}$ is the net-of-tax rate of return to wealth and g is the economy's growth rate) tends to magnify steady-state wealth inequalities. This is particularly easy to see in dynamic model with random multiplicative shocks, where the steady-state distribution of wealth has a Pareto shape, with a Pareto exponent that is directly determined by $\bar{r}-g$ (for a given structure of shocks).

15.5.4.1 An Illustrative Example with Closed-Form Formulas

To illustrate this point, consider the following model with discrete time $t=0, 1, 2, \ldots$. The model can be interpreted as an annual model (with each period lasting $H=1$ year), or a generational model (with each period lasting $H=30$ years), in which case saving tastes can be interpreted as bequest tastes. Suppose a stationary population $N_t=[0, 1]$ made of a continuum of agents of size one, so that aggregate and average variables are the same for wealth and national income: $W_t=w_t$ and $Y_t=y_t$. Effective labor input $L_t=N_t \cdot h_t=h_0 \cdot (1+g)^t$ grows at some exogenous, annual productivity rate g. Domestic output is given by some production function $Y_{dt}=F(K_t, L_t)$.

We suppose that each individual $i \in [0, 1]$ receives the same labor income $y_{Lti}=y_{Lt}$ and has the same annual rate of return $r_{ti}=r_t$. Each agent chooses c_{ti} and w_{t+1i} so as to maximize a utility function of the form $V(c_{ti}, w_{ti})=c_{ti}^{1-s_{ti}} w_{ti}^{s_{ti}}$, with wealth (or bequest) taste parameter s_{ti} and budget constraint $c_{ti}+w_{t+1i} \leq y_{Lt}+(1+r_t) \cdot w_{ti}$. Random shocks only come from idiosyncratic variations in the saving taste parameters s_{ti}, which are supposed to be drawn according to some i.i.d. random process with mean $s=E(s_{ti})<1$.[36]

With the simple Cobb–Douglas specification for the utility function, utility maximization implies that consumption c_{ti} is a fraction $1-s_{ti}$ of $y_{Lt}+(1+r_t) \cdot w_{ti}$, the total resources (income plus wealth) available at time t. Plugging this formula into the budget constraint, we have the following individual-level transition equation for wealth:

$$w_{t+1i}=s_{ti} \cdot [y_{Lt}+(1+r_t) \cdot w_{ti}] \tag{15.1}$$

At the aggregate level, since by definition national income is equal to $y_t=y_{Lt}+r_t \cdot w_t$, we have

$$w_{t+1}=s \cdot [y_{Lt}+(1+r_t) \cdot w_t]=s \cdot [y_t+w_t] \tag{15.2}$$

dividing by $y_{t+1} \approx (1+g) \cdot y_t$ and denoting $\alpha_t=r_t \cdot \beta_t$ the capital share and $(1-\alpha_t)=y_{Lt}/y_t$ the labor share, we have the following transition equation for the wealth–income ratio $\beta_t=w_t/y_t$:

[36] For a class of dynamic stochastic models with more general structures of preferences and shocks, see Piketty and Saez (2013).

$$\beta_{t+1} = s \cdot \frac{1-\alpha_t}{1+g} + s \cdot \frac{1+r_t}{1+g} \cdot \beta_t = \frac{s}{1+g} \cdot (1+\beta_t) \tag{15.3}$$

In the open-economy case, the world rate of return $r_t = r$ is given. From the above equation one can easily see that β_t converges toward a finite limit β if and only if

$$\omega = s \cdot \frac{1+r}{1+g} < 1$$

In case $\omega > 1$, then $\beta_t \to \infty$. In the long run, the economy is no longer a small open economy, and the world rate of return will have to fall so that $\omega < 1$.

In the closed-economy case, β_t always converges toward a finite limit, and the long-run rate of return r is equal to the marginal product of capital and depends negatively upon β. With a CES production function, for example, we have $r = F_K = \alpha \cdot \beta^{-1/\sigma}$ (see Section 15.5.3).

Setting $\beta_{t+1} = \beta_t$ in Equation (15.3), we obtain the steady-state wealth–income ratio:

$$\beta_t \to \beta = s/(g+1-s) = \tilde{s}/g$$

where $\tilde{s} = s(1+\beta) - \beta$ is the steady-state saving rate expressed as a fraction of national income.

Noting $z_{ti} = w_{ti}/w_t$ the normalized individual wealth, and dividing both sides of Equation (15.1) by $w_{t+1} \approx (1+g) \cdot w_t$, the individual-level transition equation for wealth can be rewritten as follows[37]:

$$z_{t+1i} = \frac{s_{ti}}{s} \cdot [(1-\omega) + \omega \cdot z_{ti}] \tag{15.4}$$

Standard convergence results (e.g., Hopehnayn and Prescott, 1992, Theorem 2, p. 1397) then imply that the distribution $\psi_t(z)$ of relative wealth will converge toward a unique steady-state distribution $\psi(z)$ with a Pareto shape and a Pareto exponent that depends on the variance of taste shocks s_{ti} and on the ω coefficient.

For instance, assume simple binomial taste shocks: $s_{ti} = s_0 = 0$ with probability $1-p$, and $s_{ti} = s_1 > 0$ with probability p (with $s = p \cdot s_1$ and $\omega < 1 < \omega/p$). The long-run distribution function $1 - \Psi_t(z) = \text{proba}(z_{ti} \geq z)$ will converge for high z toward

$$1 - \Phi(z) \approx \left(\frac{\lambda}{z}\right)^a,$$

with a constant term λ

[37] Note that $y_{Lt} = (1-\alpha) \cdot y_t$, where $\alpha = r \cdot \beta = r \cdot s/(1+g-s)$ is the long-run capital share. Note also that the individual-level transition equation given below holds only in the long run (i.e., when the aggregate wealth–income ratio has already converged).

$$\lambda = \frac{1-\omega}{\omega - p},$$

a Pareto coefficient a

$$a = \frac{\log(1/p)}{\log(\omega/p)} > 1, \tag{15.5}$$

and an inverted Pareto coefficient b

$$b = \frac{a}{a-1} = \frac{\log(1/p)}{\log(1/\omega)} > 1.$$

To see this, note that the long-run distribution with $\omega < 1 < \omega/p$ looks as follows: $z=0$ with probability $1-p$, $z = \frac{1-\omega}{p}$ with probability $(1-p)\cdot p$, …, and $z = z_k = \frac{1-\omega}{\omega - p}\cdot\left[\left(\frac{\omega}{p}\right)^k - 1\right]$ with probability $(1-p)\cdot p^k$. As $k \to +\infty$, $z_k \approx \frac{1-\omega}{\omega-p}\cdot\left(\frac{\omega}{p}\right)^k$ The cumulated distribution is given by $1-\Phi(z_k) = \text{proba}(z \geq z_k) = \sum_{k' \geq k}(1-p)\cdot p^{k'} = p^k$. It follows that as $z \to +\infty$, $\log[1-\Phi(z)] \approx a\cdot[\log(\lambda) - \log(z)]$, i.e., $1-\Phi(z) \approx (\lambda/z)^a$. In case $\omega/p < 1$, then $z_k = \frac{1-\omega}{p-\omega}\cdot\left[1-\left(\frac{\omega}{p}\right)^k\right]$ has a finite upper bound $z_1 = \frac{1-\omega}{p-\omega}$.[38]

As ω rises, a declines and b rises, which means that the steady-state distribution of wealth is more and more concentrated.[39] Intuitively, an increase in $\omega = s\cdot\frac{1+r}{1+g}$ means that the multiplicative wealth inequality effect becomes larger as compared to the equalizing labor income effect, so that steady-state wealth inequalities get amplified.

In the extreme case where $\omega \to 1^-$ (for given $p < \omega$), $a \to 1^+$ and $b \to +\infty$ (infinite inequality). That is, the multiplicative wealth inequality effect becomes infinite as compared to the equalizing labor-income effect. The same occurs as $p \to 0^+$ (for given $\omega > p$): an infinitely small group gets infinitely large random shocks.[40] Explosive wealth inequality paths can also occur in case the taste parameter s_{ti} is higher on average for individuals with high initial wealth.[41]

[38] See Piketty and Saez (2013, working paper version, pp. 51–52).

[39] A higher inverted Pareto coefficient b (or, equivalently, a lower Pareto coefficient a) implies a fatter upper tail of the distribution and higher inequality. On the historical evolution of Pareto coefficients, see Atkinson et al. (2011, pp. 13–14 and 50–58).

[40] In the binomial model, one can directly compute the "empirical" inverted Pareto coefficient $b' = \frac{E(z|z \geq z_k)}{z_k} \to \frac{1-p}{1-\omega}$ as $k \to +\infty$. Note that $b' \simeq b$ if p, $\omega \simeq 1$, but that the two coefficients generally differ because the true distribution is discrete, while the Pareto law approximation is continuous.

[41] Kuznets (1953) and Meade (1964) were particularly concerned about this potentially powerful unequalizing force.

15.5.4.2 Pareto Formulas in Multiplicative Random Shocks Models

More generally, one can show that all models with multiplicative random shocks in the wealth accumulation process give rise to distributions with Pareto upper tails, whether the shocks are binomial or multinomial, and whether they come from tastes or other factors. For instance, the shock can come from the rank of birth, such as in the primogeniture model of Stiglitz (1969),[42] or from the number of children (Cowell, 1998),[43] or from rates of return (Benhabib et al., 2011, 2013; Nirei, 2009). Whenever the transition equation for wealth can be rewritten so as take a multiplicative form

$$z_{t+1i} = \omega_{ti} \cdot z_{ti} + \varepsilon_{ti}$$

where ω_{ti} is an i.i.d. multiplicative shock with mean $\omega = E(\omega_{ti}) < 1$, and ε_{ti} an additive shock (possibly random), then the steady-state distribution has a Pareto upper tail with coefficient a, which must solve the following equation:

$$E\left(\omega_{ti}^{a}\right) = 1.$$

A special case is when $p \cdot (\omega/p)^{a} = 1$, that is $a = \log(1/p)/\log(\omega/p)$, the formula given in Equation (15.5) above. More generally, as long as $\omega_{ti} > 1$ with some positive probability, there exists a unique $a > 1$, so that $E(\omega_{ti}^{a}) = 1$. One can easily see that for a given average $\omega = E(\omega_{ti}) < 1$, $a \rightarrow 1$ (and thus wealth inequality tends to infinity) if the variance of shocks goes to infinity, and $a \rightarrow \infty$ if the variance goes to zero.

Which kind of shocks have mattered most in the historical dynamics of the distribution of wealth? Many different kinds of individual-level random shocks play an important role in practice, and it is difficult to estimate the relative importance of each of them. One robust conclusion, however, is that for a given variance of shocks, steady-state wealth concentration is always a rising function of $r - g$. That is, due to cumulative dynamic effects, relatively small changes in $r - g$ (say, from $r - g = 2\%$ per year to $r - g = 3\%$ per year) can make a huge difference in terms of long-run wealth inequality.

For instance, if we interpret each period of the discrete-time model described above as lasting H years (with $H = 30$ years = generation length), and if r and g denote instantaneous rates, then the multiplicative factor ω can be rewritten as

[42] With primogeniture (binomial shock), the formula is exactly the same as before. See, e.g., Atkinson and Harrison (1978, p. 213), who generalize the Stiglitz (1969) formula and get: $a = \log(1 + n)/\log(1 + sr)$, with s the saving rate out of capital income. This is the same formula as $a = \log(1/p)/\log(\omega/p)$: with population growth rate per generation $= 1 + n$, the probability that a good shock occurs—namely, being the eldest son—is given by $p = 1/(1 + n)$. Menchik (1980), however, provides evidence on estate division in the United States, showing that equal sharing is the rule.

[43] The Cowell result is more complicated because families with many children do not return to zero (unless infinite number of children), so there is no closed form formula for the Pareto coefficient a, which must solve the following equation: $\sum \frac{p_k \cdot k}{2} \left(\frac{2 \cdot \omega}{k}\right)^{a} = 1$, where p_k = fraction of parents who have k children, with $k = 1, 2, 3$, etc., and ω = average generational rate of wealth reproduction.

$$\omega = s \cdot \frac{1+R}{1+G} = s \cdot e^{(r-g)H}$$

with $1+R=e^{rH}$ the generational rate of return and $1+G=e^{gH}$ the generational growth rate. If $r-g$ rises from $r-g=2\%$ to $r-g=3\%$, then with $s=20\%$ and $H=30$ years, $\omega = s \cdot e^{(r-g)H}$ rises from $\omega=0.36$ to $\omega=0.49$. For a given binomial shock structure $p=10\%$, this implies that the resulting inverted Pareto coefficient $b=(\log(1/p))/(\log(1/\omega))$ shifts from $b=2.28$ to $b=3.25$. This corresponds to a shift from an economy with moderate wealth inequality (say, with a top 1% wealth share around 20–30%) to an economy with very high wealth inequality (say, with a top 1% wealth share around 50–60%).

Last, if we introduce taxation into the dynamic wealth accumulation model, then one naturally needs to replace r by the after-tax rate of return $\bar{r}=(1-\tau)\cdot r$, where τ is the equivalent comprehensive tax rate on capital income, including all taxes on both flows and stocks. That is, what matters for long-run wealth concentration is the differential $\bar{r}-g$ between the net-of-tax rate of return and the growth rate. This implies that differences in capital tax rates and tax progressivity over time and across countries can explain large differences in wealth concentration.[44]

15.5.4.3 On the Long-Run Evolution of $\bar{r}-g$

The fact that steady-state wealth inequality is a steeply increasing function of $\bar{r}-g$ can help explain some of the historical patterns analyzed in Section 15.3.

First, it is worth emphasizing that during most of history, the gap $\bar{r}-g$ was large, typically of the order of 4–5% per year. The reason is that growth rates were close to zero until the industrial revolution (typically less than 0.1–0.2% per year), while the rate of return to wealth was generally of the order of 4–5% per year, in particular for agricultural land, by far the most important asset.[45] We have plotted on Figure 15.27 the world GDP growth rates since Antiquity (computed from Maddison, 2010) and estimates of the average return to wealth (from Piketty, 2014). Tax rates were negligible prior to the twentieth century, so that after-tax rates of return were virtually identical to pretax rates of return, and the $\bar{r}-g$ gap was as large as the $r-g$ gap (Figure 15.28).

The very large $\bar{r}-g$ gap until the late nineteenth to early twentieth century is in our view the primary candidate explanation as to why the concentration of wealth has been so large during most of human history. Although the rise of growth rates from less than 0.5% per year before the eighteenth century to about 1–1.5% per year during the eighteenth to

[44] For instance, simulation results suggest that differences in top inheritance tax rates can potentially explain a large fraction of the gap in wealth concentration between countries such as Germany and France (see Dell, 2005).

[45] In traditional agrarian societies, e.g., in eighteenth-century Britain or France, the market value of agricultural land was typically around 20–25 years of annual land rent, which corresponds to a rate of return of about 4–5%. Returns on more risky assets such as financial loans were sometime much higher. See Piketty (2014).

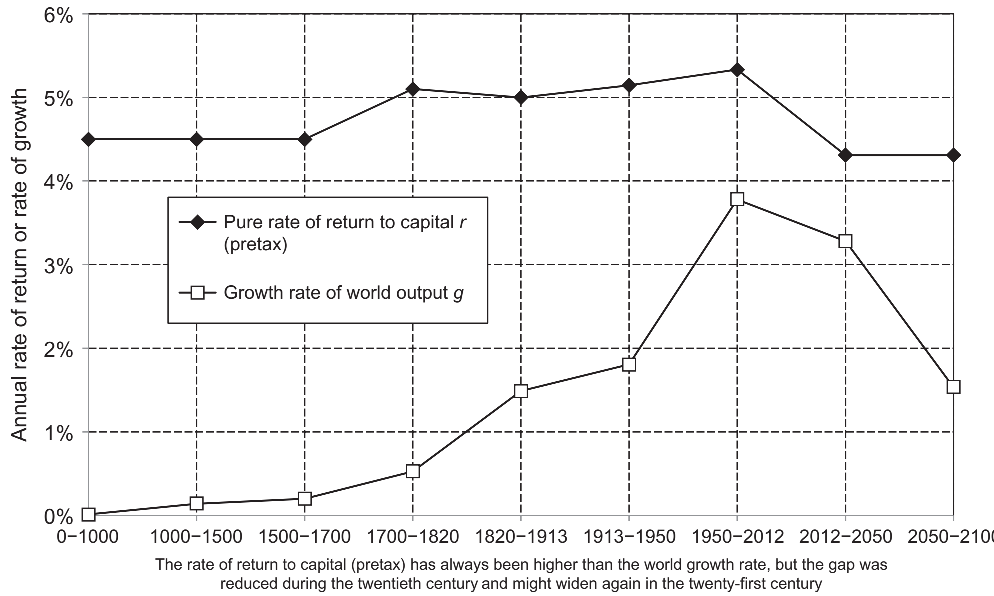


Figure 15.27 Rate of return versus growth rate at the world level, from antiquity until 2100.

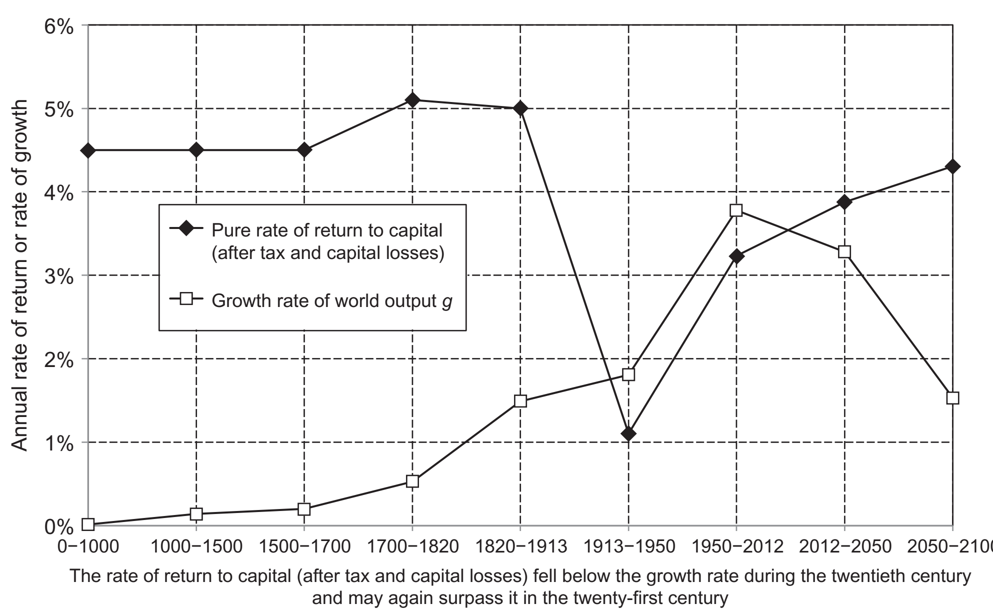


Figure 15.28 After tax rate of return versus growth rate at the world level, from antiquity until 2100.

nineteenth centuries was sufficient to make a huge difference in terms of population and living standards, it had a relatively limited impact on the $\bar{r}-g$ gap: $\bar{r}$ remained much bigger than g.[46]

The spectacular fall of the $\bar{r}-g$ gap in the course of the twentieth century can also help understand the structural decline of wealth concentration, and in particular why wealth concentration did not return to the extreme levels observed before the World Wars. The fall of the $\bar{r}-g$ gap during the twentieth century has two components: a large rise in g and a large decline in $\bar{r}$. Both, however, might well turn out to be temporary.

Start with the rise in g. The world GDP growth rate was almost 4% during the second half of the twentieth century. This is due partly to a general catch up process in per capita GDP levels (first in Europe and Japan between 1950 and 1980, and then in China and other emerging countries starting around 1980–1990), and partly to unprecedented population growth rates (which account for about half of world GDP growth rates over the past century). According to UN demographic projections, world population growth rates should sharply decline and converge to 0% during the second half of the twenty-first century. Long run per capita growth rates are notoriously difficult to predict: they might be around 1.5% per year (as posited on Figure 15.27 for the second half of the twenty-first century), but some authors—such as Gordon (2012)—believe that they could be less than 1%. In any case, it seems plausible that the exceptional growth rates of the twentieth century will not happen again—at least regarding the demographic component—and that g will indeed gradually decline during the twenty-first century.

Looking now at $\bar{r}$, we also see a spectacular decline during the twentieth century. If we take into account both the capital losses (fall in relative asset prices and physical destructions) and the rise in taxation, the net-of-tax, net-of-capital-losses rate of return $\bar{r}$ fell below the growth rate during the entire twentieth century after World War I.

Other forms of capital shocks could occur in the twenty-first century. But assuming no new shock occurs, and assuming that rising international tax competition to attract capital leads all forms of capital taxes to disappear in the course of the twenty-first century (arguably a plausible scenario, although obviously not the only possible one), the net-of-tax rate of return $\bar{r}$ will converge toward the pretax rate of return r, so that the $\bar{r}-g$ gap will again be very large in the future. Other things equal, this force could lead to rising wealth concentration during the twenty-first century.

The $\bar{r}-g$ gap was significantly larger in Europe than in the United States during the nineteenth century (due in particular to higher population growth in the New World). This fact can contribute to explain why wealth concentration was also higher in Europe. The $\bar{r}-g$ gap dramatically declined in Europe during the twentieth

[46] It is also possible that the rise of the return to capital during the eighteenth to nineteenth centuries was somewhat larger than the lower-bound estimates that we report on Figure 15.27, so that the $r-g$ gap perhaps did not decline at all. See Piketty (2014) for a more elaborate discussion.

century—substantially more than the United States, which can, in turn, explain why wealth has become structurally less concentrated than in the United States. The higher level of labor income inequality in the United States in recent decades, as well as the sharp drop in tax progressivity, also contribute to higher wealth concentration in the United States (see Saez and Zucman, 2014). Note, however, that the United States is still characterized by higher population growth (as compared to Europe and Japan), and that this tends to push in the opposite direction (i.e., less wealth concentration). So whether the wealth inequality gap with Europe will keep widening in coming decades is very much an open issue at this stage.

More generally, we should stress that although the general historical pattern of $\bar{r} - g$ (both over time and across countries) seems consistent with the evolution of wealth concentration, other factors do also certainly play an important role in wealth inequality.

One such factor is the magnitude of idiosyncratic shocks to rates of return r_{ti}, and the possibility that average rates of return $r(w) = E(r_{ti}|w_{ti} = w)$ vary with the initial wealth levels. Existing evidence on returns to university endowments suggests that larger endowments indeed tend to get substantially larger rates of returns, possibly due to scale economies in portfolio management (Piketty 2014, Chapter 12). The same pattern is found for the universe of U.S. foundations (Saez and Zucman, 2014). Evidence from *Forbes* global wealth rankings also suggests that higher wealth holders tend to get higher returns. Over the 1987–2013 period, the top fractiles (defined in proportion to world adult population) of the *Forbes* global billionaire list have been growing on average at about 6–7% per year in real terms, when average adult wealth at the global level was rising at slightly more than 2% per year (see Table 15.1).

Whatever the exact mechanism might be, this seems to indicate that the world distribution of wealth is becoming increasingly concentrated, at least at the top of the distribution. It should be stressed again, however, that available data is of relatively low quality. Little is known about how the global wealth rankings published by magazines

Table 15.1 The growth rate of top global wealth, 1987–2013

Average real growth rate per year (after deduction of inflation)	**1987–2013**
The top 1/(100 million) highest wealth holders (about 30 adults out of 3 billions in 1980s, and 45 adults out of 4.5 billions in 2010s)	6.8%
The top 1/(20 million) highest wealth holders (about 150 adults out of 3 billions in 1980s, and 225 adults out of 4.5 billions in 2010s)	6.4%
Average world wealth per adult	2.1%
Average world income per adult	1.4%
World adult population	1.9%
World GDP	3.3%

Between 1987 and 2013, the highest global wealth fractiles have grown at 6–7% per year, versus 2.1% for average world wealth and 1.4% for average world income. All growth rates are net of inflation (2.3% per year between 1987 and 2013).

are constructed, and it is likely that they suffer from various biases. They also focus on such a narrow fraction of the population that they are of limited utility for a comprehensive study of the global distribution of wealth. For instance, what happens above \$1 billion does not necessarily tell us much about what happens between \$10 and 100 million. This is a research area where a lot of progress needs to be made.

15.5.5 The Steady-State Level of the Inheritance Share: $\varphi = \varphi(g)$

15.5.5.1 The Impact of Saving Motives, Growth, and Life Expectancy

The return of high wealth–income ratios β does not necessarily imply the return of inheritance. From a purely logical standpoint, it is perfectly possible that the steady-state $\beta = s/g$ rises (say, because g goes down and s remains relatively high, as we have observed in Europe and Japan over the recent decades), but that all saving flows come from life-cycle wealth accumulation and pension funds, so that the inheritance share φ is equal to zero. Empirically, however, this does not seem to be the case. From the (imperfect) data that we have, it seems that the rise in the aggregate wealth–income ratio β has been accompanied by a rise in the inheritance share φ, at least in Europe.

This suggests that the taste for leaving bequests (and/or the other reasons for dying with positive wealth, such as precautionary motives and imperfect annuity markets) did not decline over time. Empirical evidence shows that the distribution of saving motives varies a lot across individuals. It could also be that the distribution of saving motives is partly determined by the inequality of wealth. Bequests might partly be a luxury good, in the sense that individuals with higher relative wealth also have higher bequest taste on average. Conversely, the magnitude of bequest motives has an impact on the steady-state level of wealth inequality. Take, for instance, the dynamic wealth accumulation model described above. In that model we implicitly assume that individuals leave wealth to the next generation. If they did not, the dynamic cumulative process would start at zero all over again at each generation, so that steady-state wealth inequality would tend to be smaller.

Now, assume that we take as given the distribution of bequest motives and saving parameters. Are there reasons to believe that changes in the long-run growth rate g or in the demographic parameters (such as life expectancy) can have an impact on the inheritance share φ in total wealth accumulation?

This question has been addressed by a number of authors, such as Laitner (2001) and DeLong (2003).[47] According to DeLong, the share of inheritance in total wealth accumulation should be higher in low-growth societies, because the annual volume of new savings is relatively small in such economics (so that in effect most wealth originates from inheritance). Using our notations, the inheritance share $\varphi = \varphi(g)$ is a decreasing function of the growth rate g.

[47] See also Davies et al. (2012, pp. 123–124).

This intuition is interesting (and partly correct) but incomplete. In low-growth societies, such as preindustrial societies, the annual volume of new savings—for a given aggregate β—is indeed low in steady state: $s = g \cdot \beta$. In contrast, the flow of inheritances is given by: $b_y = \mu \cdot m \cdot \beta$ (see Section 15.4). Therefore, for given μ and m, inheritance flows tend to dominate saving flows in low-growth economies, and conversely in high-growth economies.

For instance, if $\mu = 1$, $m = 2\%$, and $\beta = 600\%$, the inheritance flow is equal to $b_y = 12\%$. The inheritance flow b_y is four times bigger than the saving flow $s = 3\%$ if $g = 0.5\%$, it is equal to the saving flow $s = 12\%$ if $g = 2\%$, and it is 2.5 times smaller than the saving flow $s = 30\%$ if $g = 5\%$. Therefore—the argument goes—inherited wealth represents the bulk of aggregate wealth in low-growth, preindustrial societies; makes about half of aggregate wealth in medium-growth, mature economies; and a small fraction of aggregate wealth in high-growth, booming economies.

This intuition, however, is incomplete, for two reasons. First, as we already pointed out in Section 15.4, saving flows partly come from the return to inherited wealth, and this needs to be taken into account. Next, the μ parameter, i.e., the relative wealth of decedents, is endogenous and might well depend on the growth rate g, as well as on demographic parameters such as life expectancy and the mortality rate m. In the pure life-cycle model where agents die with zero wealth, μ is always equal to zero, and so is the inheritance share φ, independently of the growth rate g, no matter how small g is. For given (positive) bequest tastes and saving parameters, however, one can show that in steady state, $\mu = \mu(g, m)$ tends to be higher when growth rates g and mortality rates m are lower.

15.5.5.2 A Simple Benchmark Model of Aging Wealth and Endogenous μ

To see this point more clearly, it is necessary to put more demographic structure into the analysis. Here we follow a simplified version of the framework introduced by Piketty (2011).

Consider a continuous-time, overlapping-generations model with a stationary population $N_t = [0, 1]$ (zero population growth). Each individual i becomes adult at age $a = A$, has exactly one child at age $a = H$, and dies at age $a = D$. We assume away inter-vivos gifts, so that each individual inherits wealth solely when his or her parent dies, that is, at age $a = I = D - H$.

For example, if $A = 20$, $H = 30$, and $D = 60$, everybody inherits at age $I = D - H = 30$ years old. But if $D = 80$, then everybody inherits at age $I = D - H = 50$ years old.

Given that population N_t is assumed to be stationary, the (adult) mortality rate m_t is also stationary, and is simply equal to the inverse of (adult) life expectancy: $m_t = m = \frac{1}{D-A}$.[48]

[48] It is more natural to focus upon adults because minors usually have very little income or wealth (assuming that $I > A$, i.e., $D - A > H$, which is the case in modern societies).

For example, if $A = 20$ and $D = 60$, the mortality rate is $m = 1/40 = 2.5\%$. If $D = 80$, the mortality rate is $m = 1/60 = 1.7\%$. That is, in a society where life expectancy rises from 60 to 80 years old, the steady-state mortality rate among adults is reduced by a third. In the extreme case where life expectancy rises indefinitely, the steady-state mortality rate becomes increasingly small: one almost never dies.

Does this imply that the inheritance flow $b_y = \mu \cdot m \cdot \beta$ will become increasingly small in aging societies? Not necessarily: even in aging societies, one ultimately dies. Most importantly, one tends to die with higher and higher relative wealth. That is, wealth also tends to get older in aging societies, so that the decline in the mortality rate m can be compensated by a rise in relative decedent wealth μ (which, as we have seen, has been the case in France).

Assume for simplicity that all agents have on average the same uniform saving rate s on all their incomes throughout their life (reflecting their taste for bequests and other saving motives such a precautionary wealth accumulation) and a flat age-income profile (including pay-as-you-go pensions). Then one can show that the steady-state $\mu = \mu(g)$ ratio is given by the following formula:

$$\mu(g) = \frac{1 - e^{-(g - s \cdot r)(D - A)}}{1 - e^{-(g - s \cdot r)H}} = \frac{1 - e^{-(1-\alpha)g(D-A)}}{1 - e^{-(1-\alpha)gH}}.$$

With $\alpha = r \cdot \beta = r \cdot s/g =$ capital share in national income.

In other words, the relative wealth of decedents $\mu(g)$ is a decreasing function of the growth rate g (and an increasing function of the rate of return r or of the capital share α).[49] If one introduces taxes into the model, one can easily show that μ is a decreasing function of the growth rate g and an increasing function of net-of-tax rate of return $\bar{r}$ (or the net-of-tax capital share $\bar{\alpha}$).[50]

The intuition for this formula, which can be extended to more general saving models, is the following. With high growth rates, today's incomes are large as compared to past incomes, so the young generations are able to accumulate almost as much wealth as the older cohorts, in spite of the fact that the latter have already started to accumulate in the past, and in some cases have already received their bequests. Generally speaking, high growth rates g are favorable to the young generations (who are just starting to accumulate wealth, and who therefore rely entirely on the new saving flows out of current incomes), and tend to push for lower relative decedent wealth μ. High rates of return $\bar{r}$, by contrast,

[49] This steady-state formula applies both to the closed-economy and open-economy cases. The only difference is that the rate of return r is endogenously determined by the marginal product of domestic capital accumulation in the closed economy case (e.g., $r = F_K = a \cdot \beta^{-1/\sigma}$ with a CES production function), while it is a free parameter in the open economy setup (in which case the formula can be viewed as $\mu = \mu(g, r)$).

[50] With taxes, $\bar{r}$ also becomes a free parameter in the closed-economy model, so the formula should always be viewed as $\mu = \mu(g, \bar{r})$.

are more favorable to older cohorts, because this makes the wealth holdings that they have accumulated or inherited in the past grow faster, and tend to pusher for higher μ.

In the extreme case where $g \to \infty$, then $\mu \to 1$ (this directly follows from flat saving rates and age-labor income profiles).

Conversely, in the other extreme case where $g \to 0$, then $\mu \to \overline{\mu} = \frac{D-A}{H} > 1$.

It is worth noting that this maximal value $\overline{\mu}$ rises in proportion to life expectancy $D-A$ (for given generation length H). Intuitively, with $g \approx 0$ and uniform saving, most of wealth originates from inheritance, so that young agents are relatively poor until inheritance age $I = D - H$, and most of the wealth concentrates between age $D-H$ and D, so that relative decedent wealth $\mu \approx \overline{\mu} = \frac{D-A}{H}$.[51]

That is, as life expectancy $D-A$ rises, wealth gets more and more concentrated at high ages. This is true for any growth rate, and all the more for low growth rates. In aging societies, one inherits later in life,[52] but one inherits bigger amounts. With $g \approx 0$, one can see that both effects exactly compensate each other, in the sense that the steady-state inheritance flow b_y is entirely independent of life expectancy. That is, with $m = \frac{1}{D-A}$ and $\overline{\mu} = \frac{D-A}{H}$, we have $b_y = \overline{\mu} \cdot m \cdot \beta = \frac{\beta}{H}$, independently from $D-A$. For a given wealth–income ratio $\beta = 600\%$ and generation length $H = 30$ years, the steady-state annual inheritance flow is equal to $b_y = 20\%$ of national income, whether life expectancy is equal to $D = 60$ years or $D = 80$ years.

Strictly speaking, this is true only for infinitely small growth $g \approx 0$. However, by using the above formula one can see that for low growth rates (say, $g \approx 1$–1.5%) then the steady-state inheritance flow is relatively close to $b_y = \frac{\beta}{H}$ and is almost independent of life expectancy. It is only for high growth rates—above 2–3% per year—that the steady-state inheritance flow is reduced substantially.

15.5.5.3 Simulating the Benchmark Model

Available historical evidence shows that the slowdown of growth is the central economic mechanism explaining why the inheritance flow seems to be returning in the early twenty-first century to approximately the same level $b_y \approx 20\%$ as that observed during the nineteenth and early twentieth centuries.

By simulating a simple uniform-saving model for the French economy over the 1820–2010 period (starting from the observed age–wealth pattern in 1820, and using observed aggregate saving rates, growth rates, mortality rates, capital shocks and age–labor

[51] In the extreme case where young agents have zero wealth and agents above age $I = D - H$ have average wealth $\overline{w}$, then average wealth among the living is equal to $w = \frac{(D-I) \cdot \overline{w}}{D-A}$ and, so that $\overline{\mu} = \frac{\overline{w}}{w} = \frac{D-A}{H}$. See Piketty (2011), Propositions 1–3.

[52] Although in practice, this is partly undone by the rise of inter vivos gifts, as we have seen above.

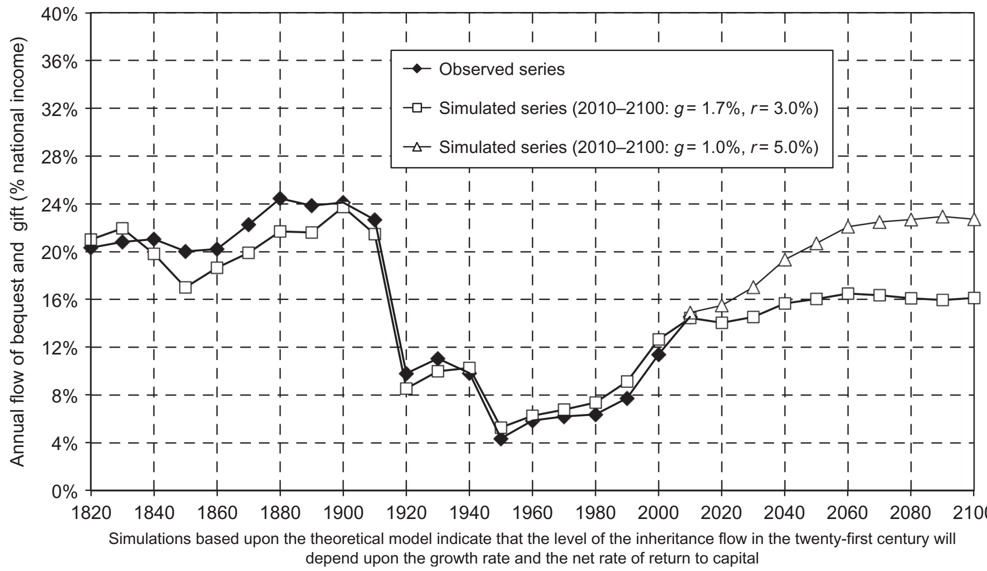


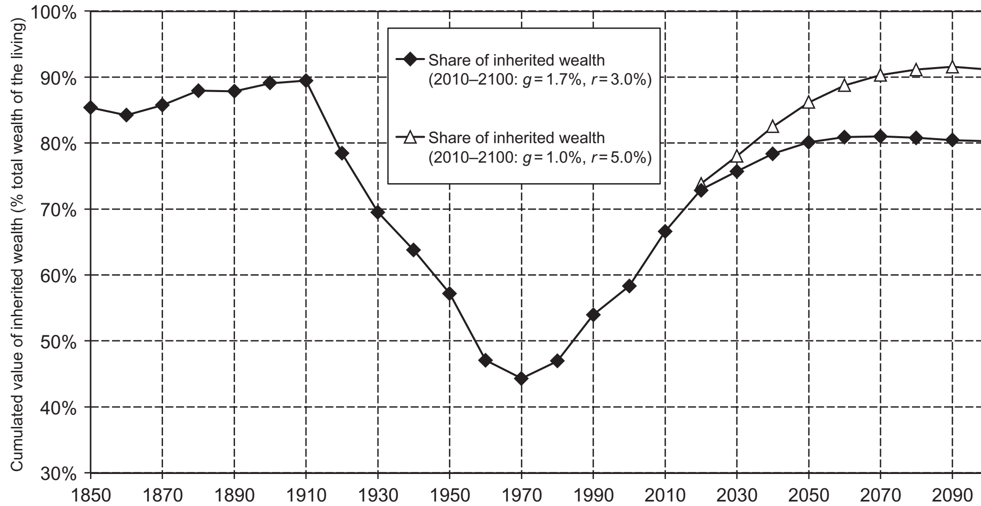
Figure 15.29 Observed and simulated inheritance flow, France 1820–2100.

income profiles over the entire period), one can reasonably well reproduce the dynamics of the age–wealth profile and hence of the μ ratio and the inheritance flow b_y over almost two centuries (see Figure 15.29).

We can then use this same model to simulate the future evolution of the inheritance flow in coming decades. As one can see on Figure 15.29, a lot depends on future values of the growth rate g and the net-of-tax rate of return $\bar{r}$ over the 2010–2100 period. Assuming $g = 1.7\%$ (which corresponds to the average growth rate observed in France between 1980 and 2010) and $\bar{r} = 3.0\%$ (which approximatively corresponds to net-of-tax average real rate of return observed in 2010), then b_y should stabilize around 16–17% in coming decades. If growth slows $g = 1.0\%$ and the net-of-tax rate of return rises to $\bar{r} = 5.0\%$ (e.g., because of a rise of the global capital share and rate of return, or because of a gradual repeal of capital taxes), b_y would keep increasing toward 22–23% over the course of the twenty-first century. The flow of inheritance would approximately return to its nineteenth and early twentieth centuries level.

In Figure 15.30, we use these projections to compute the corresponding share φ of cumulated inheritance in the aggregate wealth stock (using the PPVR definition and the same extrapolations as those described above). In the first scenario, φ stabilizes around 80%; in the second scenario, it stabilizes around 90% of aggregate wealth.

These simulations, however, are not fully satisfactory, first because a lot more data should be collected on inheritance flows in other countries, and next because one should ideally try to analyze and simulate both the flow of inheritance and the inequality of



Inherited wealth represents 80–90% of total wealth in France in the nineteenth century; this share fell to 40%–50% during the twentieth century and might return to 80%–90% during the twenty-first century

Figure 15.30 The share of inherited wealth in total wealth, France 1850–2100.

wealth. The computations presented here assume uniform saving and solely attempt to reproduce the age-average wealth profile, without taking into account within-cohort wealth inequality. This is a major limitation.

15.6. CONCLUDING COMMENTS AND RESEARCH PROSPECTS

In this chapter, we have surveyed the empirical and theoretical literature on the long-run evolution of wealth and inheritance in relation to output and income. The magnitude and concentration of wealth and inheritance (relative to national income) were very high in the eighteenth to nineteenth centuries up until World War I, then dropped precipitously during the twentieth century following World War shocks, and have been rising again in the late twentieth and early twenty-first centuries. We have showed that over a wide range of models, the long-run magnitude and concentration of wealth and inheritance are an increasing function of $\bar{r} - g$, where $\bar{r}$ is the net-of-tax rate of return to wealth and g is the economy's growth rate, and we have argued that these predictions are broadly consistent with historical patterns. These findings suggest that current trends toward rising wealth–income ratios and wealth inequality might continue during the twenty-first century, both because of the slowdown of population and productivity growth, and because of increasing international competition to attract capital.

We should stress, however, that this is an area where a lot of progress still needs to be made. Future research should particularly focus on the following issues. First, it becomes more and more important to study the dynamics of the wealth distribution from a global perspective.[53] In order to do so, it is critical to take into account existing macro data on aggregate wealth and foreign wealth holdings. Given the large movements in aggregate wealth–income ratios across countries, such macro-level variations are likely to have a strong impact on the global dynamics of the individual-level distribution of wealth. It is also critical to use existing estimates of offshore wealth and to analyze how much tax havens are likely to affect global distributional trends (see Zucman, 2014). Next, a lot more historical and international data needs to be collected on inheritance flows. Last, there is a strong need of a better articulation between empirical and theoretical research. A lot more work has yet to be done before we are able to develop rigorous and credible calibrations of dynamic theoretical models of wealth accumulation and distribution.

ACKNOWLEDGMENTS

We are grateful to the editors and to Daniel Waldenstrom for helpful comments. All series and figures presented in this chapter are available in an online data appendix.

REFERENCES

Antras, P., 2004. Is the U.S. aggregate production function Cobb-Douglas? New estimates of the elasticity of substitution. B.E. J. Macroecon. 4 (1) article 4.
Atkinson, A., 1983. The Economics of Inequality. Clarendon Press, Oxford, 330 pp.
Atkinson, A.B., 2013. Wealth and Inheritance in Britain from 1896 to the Present. Working Paper.
Atkinson, A., Harrison, A.J., 1978. Distribution of Personal Wealth in Britain, 1923–1972. Cambridge University Press, Cambridge, 330 pp.
Atkinson, A.B., Piketty, T., Saez, E., 2011. Top incomes in the long run of history. J. Econ. Lit. 49 (1), 3–71.
Benhabib, J., Bisin, A., Zhu, S., 2011. The distribution of wealth and fiscal policy in economies with finitely lived agents. Econometrica 79 (1), 123–157.
Benhabib, J., Bisin, A., Zhu, S., 2013. The Wealth Distribution in Bewley Models with Investment Risk. NYU, New York.
Blinder, A., 1988. Comments on Modigliani and Kotlikoff-Summers. In: Kessler, D., Masson, A. (Eds.), Modelling the Accumulation and Distribution of Wealth. Oxford University Press, Oxford, pp. 68–76.
Boisguillebert, P., 1695. Le détail de la France, Paris, 215 pp. Reprinted in E. Daire, Economistes financiers du 18e siècle, Paris: Guillaumin, 1843.
Bowley, A.L., 1920. The Change in the Distribution of National Income, 1880–1913. Clarendon Press, Oxford.
Champernowne, D.G., 1953. A model of income distribution. Econ. J. 63 (250), 318–351.
Chirinko, R., 2008. σ: the long and short of it. J. Macroecon. 30, 671–686.
Colquhoun, P., 1815. A Treatise on the Wealth. Power and Resources of the British Empire, London, 561 pp.
Colson, C., 1903. Cours d'économie politique. Gauthier-Villars, Paris, 1918, 1927 (several editions).
Cowell, F., 1998. Inheritance and the Distribution of Wealth. LSE, London.
Davies, J.B., Shorrocks, A.F., 1999. The distribution of wealth. In: Atkinson, A.B., Bourguignon, F. (Eds.), Handbook of Income Distribution, vol. 1. pp. 605–675 (Chapter 11).

[53] See the important pioneering work of Davies et al. (2010, 2012).

Davies, J.B., Sandström, S., Shorrocks, A., Wolff, E.N., 2010. The level and distribution of global household wealth. Econ. J. 121, 223–254.
Davies, J.B., Lluberas, R., Shorrocks, A., 2012. Global Wealth Data-book. Credit Suisse. https://www.credit-suisse.com/ch/en/news-and-expertise/research/credit-suisse-research-institute/publications.html.
Dell, F., 2005. Top incomes in Germany and Switzerland over the twentieth century. J. Eur. Econ. Assoc. 3, 412–421.
de Foville, A., 1893. The wealth of France and of other countries. J. R. Stat. Soc. 56 (4), 597–626.
DeLong, J.B., 2003. Bequests: an historical perspective. In: Munnel, A. (Ed.), The Role and Impact of Gifts and Estates. Brookings Institution, Washington, DC.
Edlund, L., Kopczuk, W., 2009. Women, wealth and mobility. Am. Econ. Rev. 99 (1), 146–178.
Fisher, I., 1920. Economists in public service. Am. Econ. Rev. 9, 5–21.
Giffen, R., 1889. The Growth of Capital. G. Bell and sons, London, 169 pp.
Goldsmith, R.W., 1985. Comparative National Balance Sheets: A Study of Twenty Countries, 1688–1978. The University of Chicago Press, Chicago, 353 pp.
Goldsmith, R.W., 1991. Pre-Modern Financial Systems. Cambridge University Press, Cambridge, 348 pp.
Gordon, R.J., 2012. Is U.S. Economic Growth Over? Faltering Innovation Confronts the Six Headwinds. NBER Working Paper 18315.
Greenwood, D., 1983. An estimation of U.S. family wealth and its distribution from microdata, 1973. Rev. Income Wealth 29, 23–43.
Helfferich, K., 1913. Deutschlands Volkswholhstand, 1888–1913. G. Stilke, Berlin.
Hopehnayn, H., Prescott, E., 1992. Stochastic monotonicity and stationary distributions for dynamic economies. Econometrica 60 (6), 1387–1406.
Jones, C.I., Romer, P.M., 2010. The new Kaldor facts: ideas, institutions, population, and human capital. Am. Econ. J. 2 (1), 224–245.
Kaldor, N., 1961. Capital accumulation and economic growth. In: Lutz, F.A., Hague, D.C. (Eds.), The Theory of Capital. Saint Martins Press, New York.
Karabarbounis, L., Neiman, B., 2014. The global decline of the labor share. Q. J. Econ. 129 (1), 61–103.
King, G., 1696. Natural and Political Observations and Conclusions Upon the State and Condition of England, 1696. pp. 29–73, 45 pp., Published in Chalmers 1804.
King, W.I., 1915. The Wealth and Income of the People of the United States. McMillan Press, New York , 278 pp.
King, W.I., 1927. Wealth distribution in the Continental United States at the close of 1921. J. Am. Stat. Assoc. 22, 135–153.
Kaplan, S.N., Rauh, J., 2013. Family, education, and sources of wealth among the richest Americans, 1982–2012. Am. Econ. Rev. 103 (3), 158–166.
Karagiannaki, E., 2011. Recent Trends in the Size and the Distribution of Inherited Wealth in the UK. London School of Economics, London, CASE Paper 146.
Kopczuk, W., Saez, E., 2004. Top wealth shares in the United States, 1916–2000: evidence from estate tax returns. Natl. Tax J. 57 (2), 445–487.
Kotlikoff, L., 1988. Intergenerational transfers and savings. J. Econ. Perspect. 2 (2), 41–58.
Kotlikoff, L., Summers, L., 1981. The role of intergenerational transfers in aggregate capital accumulation. J. Polit. Econ. 89, 706–732.
Kuznets, S., 1953. Shares of upper income groups in income and savings, 1913–1948. National Bureau of Economic Research, Cambridge, MA, 707 pp.
Lampman, R.J., 1962. The share of top wealth-holders in national wealth 1922–1956. Princeton University Press, Princeton.
Laitner, J., 1979. Bequest behaviour and the national distribution of wealth. Rev. Econ. Stud. 46 (3), 467–483.
Laitner, J., 2001. Secular changes in wealth inequality and inheritance. Econ. J. 111 (474), 691–721.
Lindert, P., 1986. Unequal English wealth since 1688. J. Polit. Econ. 94 (6), 1127–1162.
Lindert, P., 2000. Three centuries of inequality in Britain and America. In: Atkinson, A.B., Bourguignon, F. (Eds.), Handbook of Income Distribution, vol. 1. Elsevier/North Holland, Amsterdam, pp. 167–216.

Maddison, A., 2010. Historical Statistics of the World Economy: 1–2008 AD. www.ggcc.net/maddisonupdated tables.
Mallet, B., 1908. A method of estimating capital wealth from the estate duty statistics. J. R. Stat. Soc. 71 (1), 65–101.
Mallet, B., Strutt, H.C., 1915. The multiplier and capital wealth. J. R. Stat. Soc. 78 (4), 555–599.
Meade, J.E., 1964. Efficiency, Equality and the Ownership of Property. Allen and Unwin, London.
Menchik, P.L., 1980. Primogeniture, equal sharing, and the U.S. distribution of wealth. Q. J. Econ. 94 (2), 299–316.
Modigliani, F., 1986. Life cycle, individual thrift and the wealth of nations. Am. Econ. Rev. 76 (3), 297–313.
Modigliani, F., 1988. The role of intergenerational transfers and lifecycle savings in the accumulation of wealth. J. Econ. Perspect. 2 (2), 15–40.
Morigushi, C., Saez, E., 2008. The evolution of income concentration in Japan, 1886–2005: evidence from income tax statistics. Rev. Econ. Stat. 90 (4), 713–734.
Nirei, M., 2009. Pareto Distributions in Economic Growth Models. Hitotsubashi University, Tokyo, Working Paper.
Ohlsson, H., Roine, J., Waldenstrom, D., 2013. Inherited Wealth over the Path of Development: Sweden 1810-2010. Mimeo, Uppsala.
Petty, W., 1664. Verbum Sapienti. 26 pp. Published as an addendum to Petty, W., 1691. The Political Anatomy of Ireland. London, 266 pp. Reprinted in The Economic Writings of Sir William Petty, Hull, C.H. (Ed.), 2 vols., Cambridge University Press, 1989.
Piketty, T., 2011. On the long-run evolution of inheritance: France 1820–2050. Q. J. Econ. 126 (3), 1071–1131.
Piketty, T., 2014. Capital in the 21st Century. Harvard University Press, Cambridge.
Piketty, T., Saez, E., 2013. A theory of optimal inheritance taxation. Econometrica 81 (5), 1851–1886.
Piketty, T., Zucman, G., 2014. Capital is back: wealth-income ratios in rich countries 1700–2010. Q. J. Econ. 129 (3), 1255–1310.
Piketty, T., Postel-Vinay, G., Rosentha, J.-L., 2006. Wealth concentration in a developing economy: Paris and France, 1807–1994. Am. Econ. Rev. 96 (1), 236–256.
Piketty, T., Postel-Vinay, G., Rosenthal, J.-L., 2013. Inherited versus self-made wealth: theory and evidence from a rentier society (1872–1927). Explor. Econ. Hist. 51, 21–40.
Roine, J., Waldenstrom, D., 2009. Wealth concentration over the path of development: Sweden, 1873–2006. Scand. J. Econ. 111, 151–187.
Saez, E., Zucman, G., 2014. Wealth Inequality in the United States since 1913: Evidence from Capitalized Income Tax Data. Working Paper.
Schinke, C., 2013. Inheritance in Germany 1911 to 2009. PSE Working Paper.
Séaillès, J., 1910. La répartition des fortunes en France. Editions Felix Alcan, Paris, 143 pp.
Stamp, J.C., 1919. The wealth and income of the chief powers. J. R. Stat. Soc. 82 (4), 441–507.
Stewart, C., 1939. Income capitalization as a method of estimating the distribution of wealth by size group. Studies in Income and Wealth, vol. 3. National Bureau of Economic Research, New York.
Stiglitz, J.E., 1969. Distribution of income and wealth among individuals. Econometrica 37 (3), 382–397.
Strutt, H.C., 1910. Notes on the distribution of estates in France and the United Kingdom. J. R. Stat. Soc. 73 (6), 634–644.
Vauban, S., 1707. Projet d'une dixme royale. Librairie Félix Alcan, Paris, 267 pp. Reprinted in E. Daire, Economistes financiers du 18e siècle, Paris: Guillaumin, 1843.
Vaughan, R.N., 1979. Class behaviour and the distribution of wealth. Rev. Econ. Stud. 46 (3), 447–465.
Waldenstrom, D., 2009. Lifting All Boats? The Evolution of Income and Wealth Inequality over the Path of Development. PhD Dissertation, Lunds universitet, Lund.
Wolff, E., Gittleman, M., 2013. Inheritances and the distribution of wealth or whatever happened to the great inheritance boom? J. Econ. Inequal. 1–30, available online at http://dx.doi.org/10.1007/s10888-013-9261-8.
Zucman, G., 2013. The missing wealth of nations: are Europe and the U.S. net debtors or net creditors? Q. J. Econ. 128 (3), 1321–1364.
Zucman, G., 2014. Taxing Across Borders: Tracking Personal Wealth and Corporate Profits. J. Econ. Perspect. forthcoming.