Income and Wealth Inequality in America, 1949-2016 *

Moritz Kuhn† Moritz Schularick‡ Ulrike I. Steins§

June 7, 2018

Abstract: This paper introduces a new long-run dataset based on archival data from historical waves of the Survey of Consumer Finances. The household-level data allow us to study the joint distributions of household income and wealth since 1949. We expose the central importance of portfolio composition and asset prices for wealth dynamics in postwar America. Asset prices shift the wealth distribution because the composition and leverage of household portfolios differ systematically along the wealth distribution. Middle-class portfolios are dominated by housing, while rich households predominantly own equity. An important consequence is that the top and the middle of the distribution are affected differentially by changes in equity and house prices. Housing booms lead to substantial wealth gains for leveraged middle-class households and tend to decrease wealth inequality, all else equal. Stock market booms primarily boost the wealth of households at the top of the distribution. This race between the equity market and the housing market shaped wealth dynamics in postwar America and decoupled the income and wealth distribution over extended periods. The historical data also reveal that no progress has been made in reducing income and wealth inequalities between black and white households over the past 70 years, and that close to half of all American households have less wealth today in real terms than the median household had in 1970.

JEL: D31, E21, E44, N32

Keywords: Income and wealth inequality, household portfolios, historical micro data

*We thank Lukas Gehring for his outstanding research assistance during the early stages of this project. We thank participants at the NBER Summer Institute, at the wid.world conference at PSE, ASU, INET Cambridge, ASSA Philadelphia, SED Edinburgh, SAET, as well as seminar participants at Humboldt University of Berlin, DIW, Konstanz, Munich, Oslo, and the Federal Reserve Banks of St. Louis, New York, and Minneapolis. We are grateful to Christian Bayer, Jesse Bricker, Emma Enderby, Kyle Herkenhoff, Dirk Krüger, Per Krusell, Felix Kubler, Olivier Godechot, Thomas Piketty, Josep Pijoan-Mas, Ed Prescott, José-Víctor Ríos-Rull, Aysegul Sahin, Petr Sedlacek, Thomas Steger, Felipe Valencia, Gustavo Ventura, Gianluca Violante, and Gabriel Zucman for their helpful comments and suggestions. Steins gratefully acknowledges financial support from a scholarship of the Science Foundation of Sparkassen-Finanzgruppe. The usual disclaimer applies.

†University of Bonn, CEPR, and IZA, Adenauerallee 24-42, 53113 Bonn, Germany, mokuhn@uni-bonn.de

‡University of Bonn, and CEPR, Adenauerallee 24-42, 53113 Bonn, Germany, schularick@uni-bonn.de

§University of Bonn, Adenauerallee 24-42, 53113 Bonn, Germany, ulrike.steins@uni-bonn.de

1 Introduction

We live in unequal times. The causes and consequences of widening disparities in income and wealth have become a defining debate of our age. Recent studies have made major inroads into documenting trends in either income or wealth inequality in the United States. (Piketty and Saez (2003), Kopczuk, Saez, and Song (2010), Saez and Zucman (2016)), but we still know little about how the joint distributions of income and wealth evolved over the long run. This paper fills this gap.

The backbone of this study is a newly compiled dataset that builds on household-level information and spans the entire U.S. population over seven decades of postwar American history. We unearthed historical waves of the Survey of Consumer Finances (SCF) that were conducted by the Economic Behavior Program of the Survey Research Center at the University of Michigan from 1948 to 1977. In extensive data work, we linked the historical survey data to the modern SCFs that the Federal Reserve redesigned in 1983.[1] We call this new resource for inequality research the *Historical Survey of Consumer Finances* (HSCF).

The HSCF complements existing datasets for long-run inequality research that are based on income tax and social security records, but also goes beyond them in a number of important ways. Importantly, the HSCF is the first dataset that makes it possible to study the joint distributions of income and wealth over the long run. As a historical version of the SCF, it contains the same comprehensive income and balance sheet information as the modern SCFs. This means that we do not have to combine data from different sources or capitalize observed income tax data to generate wealth holdings. Moreover, the HSCF contains granular demographic information that can be used to study dimensions of inequality —such as long-run trends in racial inequality— that so far have been out of reach for research.

Our analysis speaks to the quest to generate realistic wealth dynamics in dynamic quantitative models (Benhabib and Bisin (2016), Fella and De Nardi (2017), Gabaix, Lasry, Lions, and Moll (2016), Hubmer, Krusell, and Smith (2017)). A key finding of our paper is that a channel that has attracted little scrutiny so far has played a central role in the evolution of wealth inequality in postwar America: asset price changes induce shifts in the wealth distribution because the composition and leverage of household portfolios differ systematically along the wealth distribution. While the portfolios of rich households are dominated by corporate and noncorporate equity, the portfolio of a typical middle-class household is highly concentrated in residential real estate and, at the same time, highly leveraged. These portfolio differences are persistent over time. We document this new stylized fact and expose

[1] A few studies such as Malmendier and Nagel (2011) or Herkenhoff (2013) exploited parts of these data to address specific questions, but no study has attempted to harmonize modern and historical data in a consistent way.

its consequences for the dynamics of the wealth distribution.

An important upshot is that the top and the middle of the distribution are affected differentially by changes in equity and house prices. Housing booms lead to substantial wealth gains for leveraged middle-class households and tend to decrease wealth inequality, all else equal. Stock market booms primarily boost the wealth of households at the top of the wealth distribution as their portfolios are dominated by listed and unlisted business equity. Portfolio heterogeneity thus gives rise to a race between the housing market and the stock market in shaping the wealth distribution. We show that over extended periods in postwar American history, such portfolio valuations effects have been predominant drivers of shifts in the distribution of wealth.

A second consequence of pronounced portfolio heterogeneity is that asset price movements can introduce a wedge within the evolution of the income and wealth distribution. For instance, rising asset prices can mitigate the effects that low income growth and declining savings rates have on wealth accumulation. This was prominently the case in the four decades before the financial crisis when the middle class rapidly lost ground to the top 10% with respect to income but, by and large, maintained its wealth share thanks to substantial gains in housing wealth. The HSCF data show that incomes of the top 10% more than doubled since 1971, while the incomes of middle-class households (50th to 90th percentile) increased by less than 40%, and those of households in the bottom 50% stagnated in real terms. In line with previous research, the HSCF data thus confirm a strong trend toward growing income concentration at the top (Piketty and Saez (2003); Kopczuk, Saez, and Song (2010)). However, when it comes to wealth, the picture is different. For the bottom 50%, wealth doubled between 1971 and 2007 despite zero income growth. For the middle class (50%-90%) and for the top 10%, wealth grew at approximately the same rate, rising by a factor of 2.5. As a result, wealth-to-income ratios increased most strongly for the bottom 90% of the wealth distribution. That the HSCF data reach back to the 1950s and 1960s, that is, before the income distribution started to widen substantially, makes it possible to expose these divergent trends.

Importantly, price effects account for a major part of the wealth gains of the middle class and the lower middle class. We estimate that between 1971 and 2007, the bottom 50% had wealth growth of 97% only because of price effects — essentially a doubling of wealth without any saving. Also, the upper half of the distribution registered wealth gains on an order of magnitude of 60% because of rising asset prices. For the bottom 50%, virtually all wealth growth over the 1971-2007 period came from higher asset prices. But even in the middle and at the top of the distribution, asset price induced gains accounted for close to half of total wealth growth, comparable to the contribution of savings flows. From a political economy perspective, it is conceivable that the strong wealth gains for the middle and lower middle class helped to dispel discontent about stagnant incomes. They may also help to explain the disconnect between trends in income and consumption inequality that have been the subject of some debate (Attanasio and Pistaferri (2016)).

When house prices collapsed in the 2008 crisis, the same leveraged portfolio position of the middle class brought about substantial wealth losses, while the quick rebound in stock markets boosted wealth at the top. Relative price changes between houses and equities after 2007 have produced the largest spike in wealth inequality in postwar American history. Surging postcrisis wealth inequality might in turn have contributed to the perception of sharply rising inequality in recent years.

Thanks to its demographic detail, we can also exploit the HSCF to shed new light on the long-run evolution of racial inequalities. The HSCF covers the entire postwar history of racial inequality and spans the pre- and post-civil rights eras. Importantly, as we have information on income and wealth, our paper does not complement only the recent studies of the long-run evolution of racial income inequality (Bayer and Charles (2017)); we also add an important new dimension: the HSCF data offer a window on long-run trends in racial *wealth inequality* that have so far remained unchartered territory.

We expose persistent and, in some respects, growing inequalities between black and white Americans. Income disparities today are as big as they were in the pre-civil rights era. In 1950, the income of the median white household was about twice as high as the income of the median black household. In 2016, black household income is still only half of the income of white households. The racial wealth gap is even wider and is still as large as it was in the 1950s and 1960s. The median black household persistently has less than 15% of the wealth of the median white household. We also find that the financial crisis has hit black households particularly hard and has undone the little progress that had been made in reducing the racial wealth gap during the 2000s (Wolff (2017)). The overall summary is bleak. In terms of labor market outcomes, we document that over seven decades, next to no progress has been made in closing the black-white income gap. The racial wealth gap is equally persistent and a stark fact of postwar American history. The typical black household remains poorer than 80% of white households.

Related literature: Research on inequality has become a highly active field, and our paper speaks to a large literature. Analytically, the paper is most closely related to recent contributions emphasizing the importance of returns on wealth for the wealth distribution. On the empirical side, this literature has mainly worked with European data, while our paper addresses the issues with long-run micro data for the United States. Bach, Calvet, and Sodini (2016) study administrative Swedish data. With regard to heterogeneity in returns along the wealth distribution, Fagereng, Guiso, Malacrino, and Pistaferri (2016) use administrative Norwegian tax data and document substantial heterogeneity in wealth returns and intergenerational persistence. For France, Garbinti, Goupille-Lebret, and Piketty (2017) analyze the long-run distribution of wealth as well as the role of return and savings rate differentials. In the American context, Wolff (2017) demonstrates the sensitivity of middle-class wealth to the house price collapse in the Great Recession. Kuhn and Ríos-Rull (2016) argue that housing wealth plays an important role for the wealth distribution.

With respect to data production and the emphasis on long-run trends, our paper complements the pioneering work of Piketty and Saez (2003) and Saez and Zucman (2016), as well as the work of Kopczuk, Saez, and Song (2010). Our paper also speaks to the more recent contribution of Piketty, Saez, and Zucman (2016), who combined micro data from tax records and household survey data to derive the distribution of income reported in the national accounts.[2] While their study focuses on the distribution of income growth, our paper sheds new light on the distribution of wealth growth over time. Saez and Zucman (2016) estimate the wealth distribution by capitalizing income flows from administrative data. This approach is advantageous for households at the top of the distribution that hold a significant part of their wealth in assets that generate taxable income flows. Yet many assets in middle-class portfolios do not generate taxable income flows — housing being a prime example. The HSCF provides long-run data on all sources of income (including capital and non-taxable income) as well as the entire household balance sheet with all assets (including residential real estate) and liabilities (including mortgage debt). Playing to the strength of our data, we complement Saez and Zucman (2016) by focusing on the bottom 90% of households, not on changes in inequality at the very top. Like Kopczuk (2015), we find that the top 10% income and wealth shares are similar in level and trend across different data sources.[3]

Theoretical work modeling the dynamics of wealth inequality has grown quickly. A common thread is that models based on labor income risk typically produce too little wealth concentration and cannot account for substantial shifts in wealth inequality that occur over short time horizons. Our paper speaks to recent work by Benhabib and Bisin (2016), Benhabib,

[2] Piketty, Saez, and Zucman (2016) use survey data from the Current Population Survey (CPS) to impute the distribution of transfers in terms of synthetic micro data. For income, they rely on the work done by Piketty and Saez (2003) that utilizes tax data.

[3] Work in labor economics often relies on data from the CPS. Examples are Gottschalk and Danziger (2005) and Burkhauser, Feng, and Jenkins (2009). Most relevant for our work is Burkhauser, Feng, Jenkins, and Larrimore (2012), who show that trends in income inequality derived from the CPS are similar to the inequality series based on tax data in Piketty and Saez (2003). They also provide a detailed discussion of the conceptual differences in measuring income in the tax and CPS data. The two most notable differences are incomes going to defined contribution plans that are recorded in the CPS but missed in the tax data and stock options that are not recorded in the CPS but measured in the tax data. They find that the differences are small overall.

Bisin, and Luo (2017), and Gabaix, Lasry, Lions, and Moll (2016), who discuss the importance of heterogeneous returns for the wealth distribution and its changes over time. In another recent paper, Hubmer, Krusell, and Smith (2017) use variants of incomplete market models to quantify the contribution of different drivers for rising wealth inequality and point to return differences and portfolio differences as a neglected line of research. Our findings support an emphasis on asset returns.[4] Glover, Heathcote, Krueger, and Rıos-Rull (2017) quantify the welfare effects of wealth changes resulting from portfolio differences and asset price changes during the Great Recession. Fella and De Nardi (2017) survey the existing literature and discuss different models from the canonical incomplete market model to models with intergenerational transmission of financial and human capital, rate of return risk on financial investments, and more sophisticated earnings dynamics.

Outline: The paper is divided into three parts. The first part documents the extensive data work that we have undertaken over the past years to construct the HSCF and to make the historical and modern SCFs consistent. The second part then exploits the new data and presents new stylized facts for long-run trends in income and wealth inequality, including racial inequalities, that emerge from the HSCF. The third part studies the joint distributions of income and wealth and exposes the central importance of asset price changes for the dynamics of the wealth distribution in postwar America. The last section concludes.

2 The Historical Survey of Consumer Finances

The SCF is a key resource for research on household finances in the United States. It is a triennial survey, and the post-1983 data are available on the website of the board of Governors of the Federal Reserve System[5]. Yet the first consumer finance surveys were conducted as far back as 1948. The early SCF waves were directed by the Economic Behavior Program of the Survey Research Center of the Institute for Social Research at the University of Michigan. The surveys were taken annually between 1948 and 1971, and then again in 1977. The raw data are kept at the Inter-University Consortium for Political and Social Research (ICPSR) at the Institute for Social Research in Ann Arbor, Michigan. Figure 1 shows an example of a page from the survey codebook in the year 1949. This section describes the dataset and documents how we linked the historical waves of the SCF to their modern counterparts. In the analysis, we use all data and abstain from any sample selection. We adjust all data for inflation using the consumer price index (CPI) and report results in 2016 dollars.

[4]See also Castaneda, Díaz-Giménez, and Ríos-Rull (2003) for a benchmark model of cross-sectional income and wealth inequality and Kaymak and Poschke (2016) for another recent attempt to explain time trends.

[5]`https://www.federalreserve.gov`

Figure 1: Example of Survey of Consumer Finances codebook from 1949

Project # 42 -4- Card III

Col. No.

23-27 Income from wages and salaries: (Add amounts entered after questions 33, 34, 35) (in farm schedule, item 44a)

Code the amount in dollars

00000. No income from wages and salaries
Y0000. Income from wages and salaries exceeds $99,999
X0000. Income from wages and salaries not ascertained (code here if Schedule II contains only a total at the bottom of the page)

28 - Income from wages and salaries, in class intervals:

1. $1-$499
2. $500-$999
3. $1,000-$1,999
4. $2,000-$2,999
5. $3,000-$3,999
6. $4,000-$4,999
7. $5,000-$7,499
8. $7,500-$9,999
9. $10,000 and over

0. No income from wages and salaries

X. Income from wages and salaries not ascertained

29 Did you (R and SU) receive any money from interest, dividends, rents, trust fund, or royalties? (Question 37) (Farm Schedule 44b)

1. Yes, recieved income from this source; less than $100
2. Yes, received income from this source; $100-499
3. Yes, received income from this source; $500-1,999
4. Yes, received income from this source; $2,000-4999
5. " " " " " " ; $5000 or over
6 " " " " " " ; Amount not ascertained
0. No, did not receive income from this source

X. Not ascertained whether received income from this source

30-34 Income from interest, dividends, royalties, rents, trust funds, business, professional practice: (Add amounts entered after questions 37, 39, 40, 41, 43 minus 42; Farm Schedule 44b)

Code the amount in dollars

00000. No income from these sources
Y0000. Income from these sources larger than $99,999*
X0000. Income from these sources not ascertained
XY000. Negative income *

The HSCF complements existing datasets for long-run trends in U.S. income and wealth inequality that Piketty and Saez (2003) and Saez and Zucman (2016) have compiled based on administrative tax data. For future researchers, it is important to have a good understanding of the relative strengths and weaknesses. A key advantage of the tax data is their compulsory collection process resulting in near-universal coverage at the top of the distribution. By contrast, survey data have to cope with nonresponses of rich households. Bricker, Henriques, Krimmel, and Sabelhaus (2016) recently argued that the latest survey methodology is so advanced that survey data provide an accurate picture even of the richest U.S. households, but some questions clearly remain.

The strength of the administrative data in terms of coverage at the top of the distribution has to be weighed against the strengths of survey data in other respects. Most importantly, the survey data contain direct measurements of assets and debt plus a whole list of additional information that makes it possible to stratify the data by demographic characteristics. The survey data also cover people who do not file taxes, and the unit of analysis is the household, not the tax unit. This structure is in line with economic models in which the household is the relevant unit for risk and resource sharing. In 2012, there were about one-third more tax units (160.7 million) than households (121.1 million) in the United States.[6]

Moreover, specific challenges arise when income tax data are used to construct wealth estimates. The capitalization method of Saez and Zucman (2016) relies on observable income tax flows that are capitalized to back out aggregate wealth positions. While ingenious as an approach, some gaps remain because a substantial part of wealth does not generate taxable income flows and has to be imputed (often on the basis of survey data). The key asset here is owner-occupied housing as well as its corresponding liability, mortgage debt. Pension assets also do not generate taxable income flows, and unrealized capital gains do not show up on tax returns until they are realized.

In the estimates of Saez and Zucman (2016), about 90% of the total wealth outside the top 10% has to be imputed. And even for the top 10%, the share of imputed wealth stands at 40%. Saez and Zucman (2016) correctly stress that the exact distribution of these assets is of minor importance for the very top of the wealth distribution. Yet for researchers interested in distributional changes outside the very top, these imputations can be binding limitations that the HSCF overcomes. The capitalization method also has to apply a uniform return within asset classes, derived from a combination of tax income data and aggregate estimates

[6]Bricker, Henriques, Krimmel, and Sabelhaus (2016) argue that relying on tax units could lead to higher measured income concentration toward the top of the distribution. The unit of analysis in the SCF is the primary economic unit (PEU) that contains persons in a household who share finances. The SCF sampling weights are constructed to be representative of all U.S. households, following the household definition of the U.S. Census Bureau. The Census household definition deviates slightly from that of a PEU as it groups people living together in a housing unit. In some cases, this definition may include several PEUs living together. Although the two concepts will lead to identical units of observation in the vast majority of cases, Kuhn and Ríos-Rull (2016) report that in 2013 the average SCF household is slightly smaller than a Census household (see also Bricker, Henriques, Krimmel, and Sabelhaus (2016)).

from the flow of funds. Kopczuk (2015) provides an illustration of how this method can lead to an upward bias of wealth concentration during low interest rate periods. Bricker, Henriques, and Hansen (2018) quantify this upward bias.

Overall, it is important to stress the complementarity of the different approaches and datasets. Researchers interested in the very top of the distribution might well prefer the administrative data, while those interested in wider groups might opt for the HSCF. Depending on the research question at hand, users of the data will have to carefully weigh the advantages of both.

2.1 Variables

The variables covered in the historical surveys correspond to those in the contemporary SCF, but the exact wording of the questions can differ from survey to survey. Financial innovations affect continuous coverage of variables across the various surveys. For instance, data on credit card balances become available after their introduction and proliferation. However, the appearance of new financial products such as credit cards does not impair the construction of consistent data series. Implicitly, these financial products are counted as zero for years before their appearance. Some variables are not continuously covered, so we have to impute values in some years. We explain the imputation procedure in the following section. Our analysis focuses on the four variables that are of particular importance for household finances: income, assets, debt, and wealth.

Income: We construct total income as the sum of wages and salaries, income from professional practice and self-employment, rental income, interest, dividends, transfer payments, as well as business and farm income. Income variables are available for all years.

Assets: The historical SCF waves contain detailed information on household assets. We group assets into the following categories: liquid assets, housing, bonds, stocks and business equity, mutual funds, the cash value of life insurance, other real estate, and cars. The coverage is comprehensive for liquid assets and housing. Liquid assets comprise the sum of checking, savings, call/money market accounts, and certificates of deposits. Information on liquid assets is available for almost every year of the dataset, except for 1964 and 1966. Data on defined contribution pensions are only available from 1983 onward. However, according to the flow of funds accounts (FFA), this variable makes up a small part of household wealth before the 1980s, so missing information before 1983 is unlikely to change the picture meaningfully.[7] The current value of cars is available in the historical files for 1955, 1956,

[7]Up to 1970, defined contribution plans correspond to less than 1% of average household wealth. Until 1977, this share increases to 1.7%.

1960, and 1967. We impute the value in other years using information on age, model, and size of the car.[8] Table 2 outlines the years and variables for when imputation is used.

Debt: Total debt consists of housing and nonhousing debt. Housing debt is calculated as the sum of debt on owner-occupied homes and debt on other real estate. All surveys except those of 1952, 1961, and 1977 include explicit information on housing debt. For 1977, only the origination value (instead of the current value) of mortgages is available. Using information on the year the mortgage was taken out, remaining maturity, and an estimated annual interest rate, we create a proxy for debt on homes for 1977.[9] All debt other than housing debt refers to and includes car loans, education loans, and other consumer loans.

Wealth: We construct wealth as the consolidated value of the household balance sheet by subtracting debt from assets. Wealth constitutes households' net worth.

2.2 Weights and imputations

The SCF is designed to be representative of the U.S. population. Yet capturing the top of the income and wealth distribution is a challenge for most surveys. The modern SCF applies a two-frame sampling scheme to oversample wealthy households. In addition to the adequate coverage of wealthy households in the historical surveys, we also need to ensure representative coverage of demographic characteristics such as race, age, and education. In the following section, we explain how we constructed the HSCF to meet these criteria.

Oversampling of wealthy households: Since its redesign in 1983, the SCF consists of two samples. The first sample is drawn using area probability sampling of the entire U.S. population based on Census information. In addition, a second so-called *list sample* is drawn based on tax information. Tax information is used to identify households that are likely to be at the top of the wealth distribution.[10] For both samples, survey weights are constructed separately. In the list sample, survey weights have to be overproportionally adjusted for nonresponses. The weight of each household corresponds to the number of

[8] Surveys up to 1971 include information on age, model, and size of the car a households owns. If a household bought a car during the previous year, the purchasing price of this car is also available. We impute the car value using the average purchasing price of cars bought in the previous year that are of the same age, size, and model. In 1977, only information on the original purchasing price and the age of the car is given. For this year, we construct the car value assuming a 10% annual depreciation rate.

[9] The surveys of 1952, 1956, 1960-1967, and 1971 contain no information on the debt on non-owner-occupied real estate. While the overall amounts tend to be small, this may reduce the debt of rich households in early survey years as they are more likely to have debt from other real estate.

[10] As tax data only provides information on income, a wealth index is constructed by capitalizing the income positions. Asset positions are estimated by dividing each source of capital income with the average rate of return of the corresponding asset.

similar households in the population. In a final step, both samples are combined and survey weights are adjusted so that the combined sample is representative of the U.S. population (see Kennickell, Woodburn, and McManus (1996)).[11] This two-frame sampling scheme yields a representative coverage of the entire population including wealthy households.

Before 1983, the HSCF sample is not supplemented by a second list sample. As a consequence, nonresponses of wealthy households are likely to be more frequent. This could lead to an underrepresentation of rich households in the historical data. We use information from the 1983 list sample to adjust for the possibility of an underrepresentation of rich households in the pre-1983 data. In a first step, we determine the proportion of households in the list sample relative to all households. Their share corresponds to approximately 2%. In a second step, we determine where the households from the list sample are located on the income and wealth distribution. We find that most observations are among the top 5% of the income and wealth distribution. Using this information, we adjust survey weights in all surveys before 1983 in two steps. First, for each year we extract all observations that are simultaneously in the top 5% of the income and wealth distribution. Secondly, we increase the weighting of these households in such a way that we effectively add 2% of wealthy households to the sample. We adjust the remaining weights accordingly. This approach is similar in spirit to Bricker, Henriques, and Hansen (2018), who adjust SCF weights inversely proportional to the overlap of the SCF sample with the Forbes list.

A concern with this adjustment could be that it relies on information from a single sample year in 1983 as list sample information is not available for other years. However, the 1962 Survey of Financial Characteristics of Consumers (SFCC) sample used a two-frame sampling scheme similar to the 1983 survey with a sample of rich households that was selected based on tax records. In Table 1, we show the share of households in the two surveys from the list sample to describe the nonresponse patterns at the top of the income and wealth distribution. There is no evidence for a time trend in nonresponses of wealthy households. Moreover, in section 3.2 we also compare the top income shares from the HSCF with the tax data and show that the weight adjustment does not produce any unusual breaks in the time series.[12]

[11] The adjustment is done by sorting all households into subgroups according to their gross asset holdings. Each subgroup may contain households from the first and second sample. Within each subgroup, the weights of households from the first and second sample are then adjusted depending on how many U.S. households they represent. If N_1 and N_2 are the number of weighted households of sample 1 and 2, respectively, then n_1 and n_2 are the number of unweighted households. The W_1 and W_2 weights are constructed for each sample separately. The adjusted weights for the combined samples, W_{12}, are then given by $W_{12} = \frac{n_i}{N_i} \frac{1}{\frac{n_1}{N_1} + \frac{n_2}{N_2}}$ for $i = 1, 2$. The fewer households an observation represents, the higher is $\frac{n_i}{N_i}$ and the more the original weight W_i is adjusted upward.

[12] As a proof of concept, we also apply in section A.1 of the appendix the adjustment to the 1983 data itself after dropping the list sample. We find that the adjustment works well for the top 10% but deteriorates toward the very right tail of the distribution. However, the very right tail of the distribution has been

Table 1: Share of respondents from list sample at the top of the distribution

	Income			Wealth		
	top 10%	top 5%	top 1%	top 10%	top 5%	top 1%
SFCC 1962	21 %	35 %	63 %	20 %	28 %	48 %
SCF 1983	17 %	34 %	88 %	17 %	32 %	72 %

Notes: Share of respondents from list sample in different parts of the income and wealth distribution. The left panel shows shares in the top of the income distribution in the 1983 SCF and the 1963 SFCC data. The right panel shows shares in the top of the wealth distribution in the 1983 SCF and the 1963 SFCC data. The shares are computed using weighted observations.

Demographic characteristics: We compare the demographic characteristics in the surveys before 1983 with data from the U.S. Census from 1940 to 1990. To obtain samples that match the Census data, we subdivide both the Census and the HSCF data into demographic subgroups. Subgroups are determined by age of the household head, college education, and race. We adjust HSCF weights by minimizing the difference between the share of each subgroup in the HSCF and the respective share in the Census.[13] As Census data are only available on a decennial basis, we linearly interpolate values between the dates.[14] In addition to these demographic characteristics, we include homeownership as an additional dimension to be matched.

Figure 2 shows the shares of 10-year age groups, college households, and black households in the Census (black squares) and in the HSCF with the adjustment of survey weights (gray dots). Using adjusted weights, the distributions of age, education, and race closely match the Census data. We match the homeownership rate equally well after the adjustment (see Figure A.1).

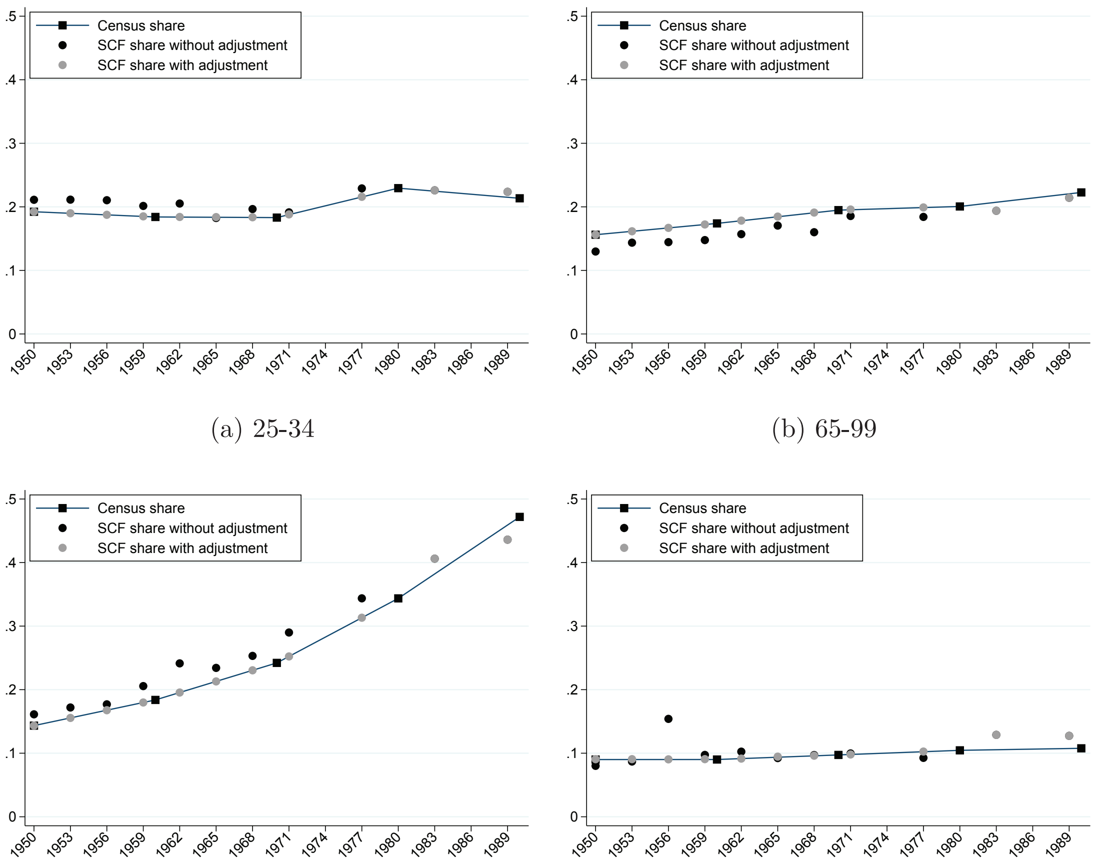
Missing variables: The imputation of missing variables is done by predictive mean matching as described in Schenker and Taylor (1996). This multiple imputation method assigns variable values by finding observations that are closest to the respective missing observations. We impute five values for each missing observation. A detailed description of the imputation method is provided in Appendix A.2. In addition, we account for a potential undercoverage of business wealth before 1983 and follow the method proposed by Saez and Zucman (2016) to adjust the observed holdings in the micro data with information from the FFA. We rely

extensively studied with tax data and is not the focus of our study.

[13]Similar to the adjustment of weights done previously, we calculate factors for each subgroup. By multiplying observations with the respective factor of their subgroup, the share of each group in the HSCF corresponds to the respective share in the Census.

[14]The distributions of demographic characteristics such as age, education, and race change gradually over time; hence, linear interpolation provides a good approximation.

Figure 2: Shares of 10-year age groups, college households, and black households in the population



(a) 25-34 (b) 65-99

(c) college (d) black

Notes: The large black squares refer to the share of the respective demographic group in the Census data. Census data are linearly interpolated in between years. The small black dots are the shares of the respective group using the original survey data. The small gray dots are the shares using the adjusted survey data. Horizontal axes show calendar time and vertical axes population shares.

on data from the 1983 and 1989 surveys and adjust business wealth and stock holdings in the earlier surveys so that the ratio of business wealth and stocks matches the 1983 and 1989 values.[15] Table 2 details the variables and their coverage, as well as the years in which we imputed data. An *O* in the table indicates that original information for the variable is

[15] Let X_{it} be business wealth or stocks of observation i in period t. The variable $\bar{\mathrm{X}}_t$ is the respective mean in period t, and X_t^{FFA} is the corresponding FFA position per household in t. The adjusted values of business wealth and stocks are then calculated as follows

$$\mathrm{X}_{it}^{adj} = \mathrm{X}_{it} \frac{\mathrm{X}_t^{FFA}}{\bar{\mathrm{X}}_t} \frac{\bar{\mathrm{X}}_{1983,1989}}{\mathrm{X}_{1983,1989}^{FFA}}$$

available for the year. An I signifies that observations for this variable were imputed. If a variable is missing in a year, we report the years of adjacent surveys that are used for the imputation in Tables A to E of the online appendix.[16]

We refer to the final dataset as the *Historical Survey of Consumer Finances* (HSCF) data. It comprises 35 survey years with cross-sectional data, totaling 110,497 household observations with demographic information and 13 continuously covered financial variables. The number of observations varies from a minimum of 1,327 in 1971 to a maximum of 6,482 in 2010. For our empirical analysis, throughout we pool the annual surveys until 1971 in three-year windows. Table A.2 in the appendix reports the number of observations for all survey years and how we pool annual surveys.

2.3 Aggregate trends

Before looking in detail at the evolution of the income and wealth distributions since World War II, the first step is to benchmark aggregate trends from the HSCF to the national income and product accounts (NIPA) and the FFA. Even high-quality micro data do not always correspond one-to-one to aggregate data as measurement concepts differ between micro surveys and national account data.[17] Yet despite the conceptual differences in measuring income and wealth, we will see that the HSCF data closely match the aggregate data.

Figure 3 compares income and wealth of the HSCF with the corresponding NIPA and FFA values.[18] FFA wealth data are calculated following Henriques and Hsu (2013), who construct wealth from the FFA to be comparable to the SCF.[19] The base period for comparisons is

[16]We exclude the survey years 1948, 1952, 1961, 1964 and 1966 because we lack information on housing, mortgages, and liquid assets. These three wealth components are held by a large fraction of households but can only be poorly inferred from information on other variables (see R^2 in Tables B, D, and E of the online appendix).

[17]For instance, Heathcote, Perri, and Violante (2010) discuss that data from the NIPA and CPS differ substantially. Indirect capital income from pension plans, nonprofit organizations, and fiduciaries, as well as employer contributions for employee and health insurance funds, are measured in the NIPA but not in household surveys such as the CPS or the SCF. In the FFA, several wealth components of the household sector are measured as residuals obtained by subtracting the positions of all other sectors from the economy-wide total (see Antoniewicz (1996), Henriques and Hsu (2013)). These residuals contain asset positions held by nonprofit organizations as well as domestic hedge funds that are not included in the SCF. Antoniewicz (1996) thoroughly discusses the measurement concepts in the SCF and FFA and concludes that there are reasons for measurement error in both datasets.

[18]Income components of the NIPA that are included are wages and salaries, proprietors' income, rental income, personal income receipts, social security, unemployment insurance, veterans' benefits, other transfers, and other net current transfer receipts from a business.

[19]This means that defined-benefit pension plans are excluded since these are not measured in the SCF and asset positions of nonprofit organizations are subtracted when possible (e.g., information on housing is provided separately for the household sector and nonprofit organizations). In addition, only mortgages and consumer credit are included as FFA debt components. However, the main adjustment to the SCF is that

Table 2: Data availability

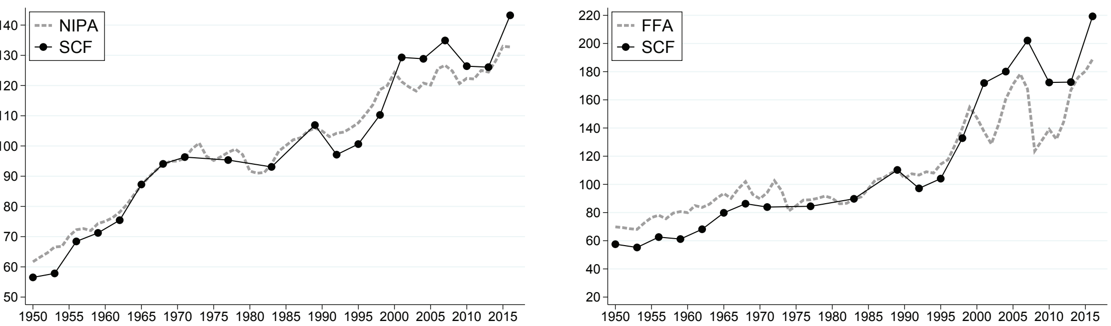
	income			financial assets			nonfinancial assets			debt			
Survey year	total	labor	labor + business	liquid assets	bonds	equity	housing	other real estate	business	total	housing	other real estate	nonhousing
1949	O	O	O	O	O	O	O	O	I	O	O	O	O
1950	O	O	O	O	O	O	O	O	O	O	O	O	O
1951	O	O	O	O	O	I	O	I	I	O	O	O	O
1952	O	O	O	O	O	O	I	O	O	I	I	I	O
1953	O	O	O	O	O	O	O	O	O	O	O	O	O
1954	O	O	O	O	O	I	O	I	I	O	O	O	O
1955	O	O	O	O	O	O	O	I	I	O	O	O	O
1956	O	O	O	O	O	I	O	I	I	I	O	I	O
1957	O	O	O	O	O	I	O	I	I	O	O	O	O
1958	O	O	O	O	O	I	O	I	I	O	O	O	O
1959	O	O	O	O	O	I	O	I	I	O	O	O	O
1960	O	I	O	O	O	O	O	O	O	I	O	I	O
1961	O	I	O	O	O	I	I	I	I	I	I	I	O
1962	O	I	O	O	O	O	O	O	O	I	O	I	O
1963	O	I	O	O	O	O	O	O	O	I	O	I	O
1964	O	I	O	I	I	O	O	I	I	I	O	I	O
1965	O	I	O	O	O	I	O	I	I	I	O	I	O
1966	O	O	O	I	I	I	O	I	I	I	O	I	I
1967	O	O	O	O	O	O	O	I	I	I	O	I	O
1968	O	O	O	O	O	O	O	O	I	O	O	O	O
1969	O	O	O	O	O	O	O	O	I	O	O	O	O
1970	O	O	O	O	O	O	O	O	O	O	O	O	O
1971	O	O	I	O	I	I	O	I	I	I	O	I	O
1977	O	O	I	O	O	O	O	O	I	O	O	O	O
1983	O	O	O	O	O	O	O	O	O	O	O	O	O
1989	O	O	O	O	O	O	O	O	O	O	O	O	O
1992	O	O	O	O	O	O	O	O	O	O	O	O	O
1995	O	O	O	O	O	O	O	O	O	O	O	O	O
1998	O	O	O	O	O	O	O	O	O	O	O	O	O
2001	O	O	O	O	O	O	O	O	O	O	O	O	O
2004	O	O	O	O	O	O	O	O	O	O	O	O	O
2007	O	O	O	O	O	O	O	O	O	O	O	O	O
2010	O	O	O	O	O	O	O	O	O	O	O	O	O
2013	O	O	O	O	O	O	O	O	O	O	O	O	O
2016	O	O	O	O	O	O	O	O	O	O	O	O	O

Notes: Data availability for different survey years. The first column shows the survey year. Each column refers to one variable in the HSCF data. The letter *O* indicates that original observations of this variables are used (i.e., no imputed observations). The letter *I* indicates that observations of this variable are imputed.

nonresidential real estate is excluded from 1989 onward (no distinction is available before 1989).

1983 to 1989 as these are the first surveys that incorporate the oversampling of wealthy households.

Figure 3: HSCF, NIPA, and FFA: income and wealth



(a) Income (b) Wealth

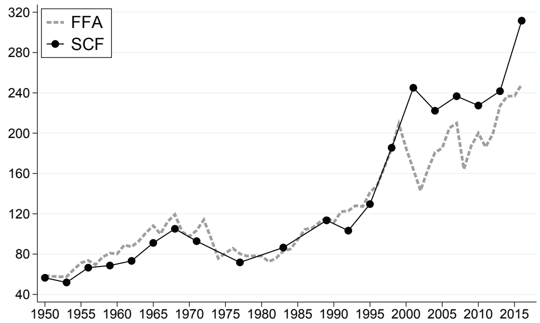
Notes: Income and wealth data from HSCF in comparison to income data from NIPA and wealth data from FFA. All data have been indexed to the 1983-1989 period (= 100). HSCF data are shown as black lines with circles, NIPA and FFA data as a gray dashed line. For the indexing period, HSCF data correspond to 80% of NIPA income and 118% of FFA wealth.

For the base period of 1983-1989, the HSCF matches 84% of income from the NIPA. Figure 3 shows that the trend in income is very similar for HSCF and NIPA data throughout the 1949-2016 time period. Looking at wealth, the trends differ only slightly. Before 1983, wealth in the HSCF is below that of the FFA. From 1983 to 1998, the two measures are about the same, and from then onward the HSCF is somewhat higher. Both wealth measures show an upward trend over time, but the increase is somewhat steeper in the HSCF.

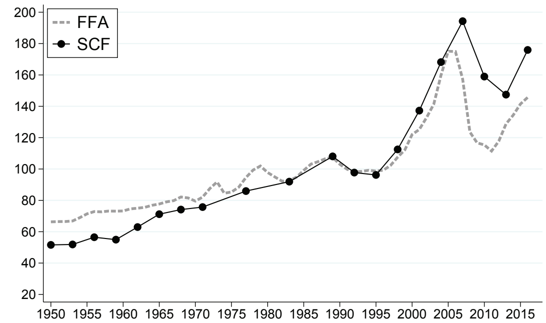
To evaluate which asset and debt positions generate the divergence in wealth estimates, Figure 4 shows different asset and debt components from the household balance sheet. Figure 4a shows financial assets. Financial assets in the HSCF increase more strongly in the 1990s than the corresponding FFA values. This difference is mainly due to distinct trends in corporate equity during the stock market boom in the second half of the 1990s. Figure 4b shows that trends for nonfinancial assets are similar in the micro and macro data. One reason for the close alignment can be seen in Figure 4c which shows that housing as the most important nonfinancial asset is covered well in the survey data. Debt is the household balance sheet component for which the HSCF matches the aggregate best, as shown in Figure 4d. The dominant component for both data sources is housing debt (Figure 4e). With respect to nonhousing debt (Figure 4f), the SCF data show somewhat lower values in the early years than the FFA but in general show a similar trend. However, nonhousing debt represents a relatively small share of total household debt.

Summing up, the HSCF matches aggregate trends of NIPA data and FFA asset and debt

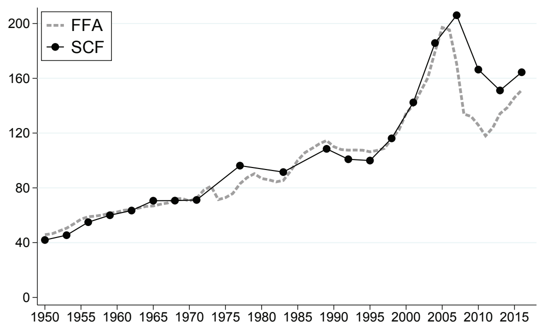
Figure 4: HSCF, NIPA, and FFA: financial and nonfinancial assets



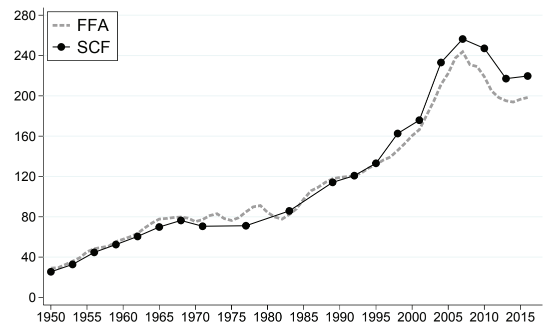
(a) Financial assets



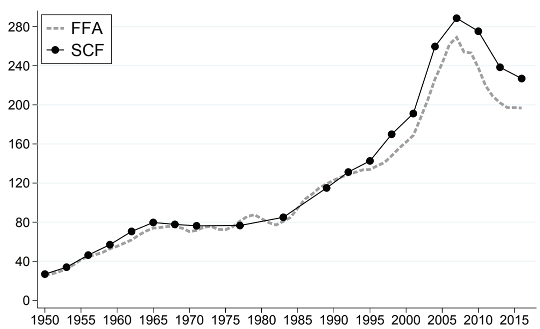
(b) Nonfinancial assets



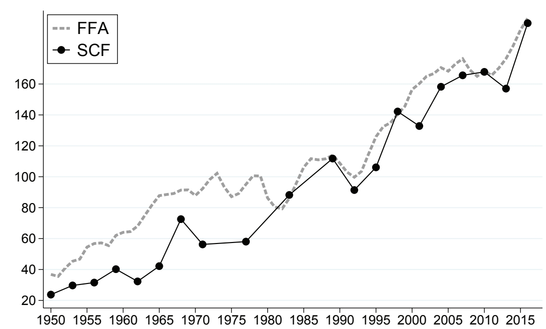
(c) Housing



(d) Total debt



(e) Housing debt



(f) Nonhousing debt

Notes: Asset and debt components of household balance sheets from the HSCF and the FFA. All data have been indexed to the 1983-1989 period (= 100). HSCF data are shown as black lines with circles, FFA data as a gray dashed line. For the indexed period, HSCF data correspond to 80% of financial assets, 137% of nonfinancial assets, 98% of housing, 86% of total debt, 93% of housing debt, and 70% of nonhousing debt.

positions. In particular, the HSCF data and the FFA show similar trends for the important categories of housing wealth and mortgage debt. Some gaps remain for financial assets such as corporate and noncorporate equity, but this is true for both the historical and post-1983 SCF data and points to conceptual differences in measurement rather than specific problems in the historical data.

3 Income and wealth inequality in the HSCF

This section presents new stylized facts for long-run trends in income and wealth inequality that the HSCF data expose. We begin by documenting the evolution of Gini coefficients for income and wealth and then turn to income and wealth shares of different groups, with a particular focus on the bottom 90%. We also use the demographic information in the HSCF data to analyze the role of demographic factors in distributional change. Importantly, we also present novel evidence on long-run trends in inequalities in income and wealth between black and white Americans.

3.1 Gini coefficients

The Gini coefficient is a comprehensive summary measure of inequality along the entire distribution. Table 3 reports Gini coefficients for income and wealth at selected points in time. The first row reports the Gini coefficient for all households; the other rows exclude the top 1% and the top 10%, respectively. We also report the full time series in Table G in the appendix.

Table 3: Gini coefficient ($\times$100) for income and wealth

		1950	1971	1989	2007	2016
income	all	44	43	53	55	58
	bottom 99%	39	39	46	47	49
	bottom 90%	31	33	39	38	39
wealth	all	81	80	79	82	86
	bottom 99%	74	74	72	74	79
	bottom 90%	61	62	63	63	70

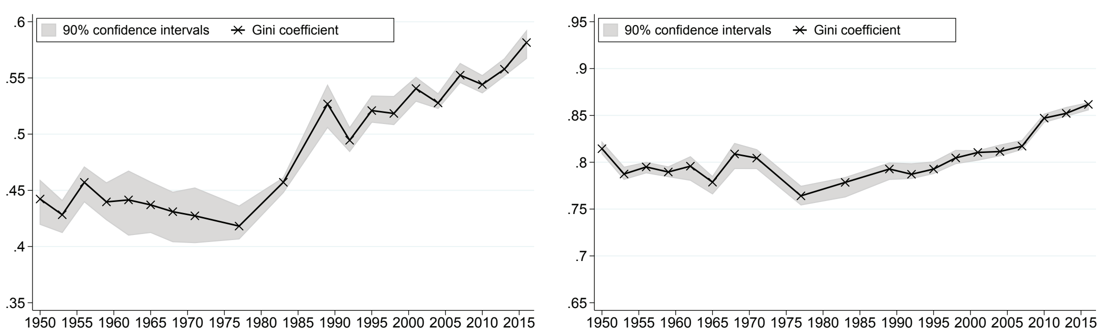
The Gini coefficients show that income and wealth inequality has increased not only across the entire population (across all households) but also among the bottom 99% and bottom

90% of households. The overall income Gini has risen from its postwar low of 0.43 in 1971 to 0.58 in 2016. Unsurprisingly, there is a substantial drop in inequality once the top 1% of the distribution is excluded, but the increase in the Gini coefficient among the bottom 99% is still substantial. Also, within the bottom 90% income inequality has widened, yet this has mainly occurred between 1971 and 1989. The rise in inequality in the past three decades has played out mainly at the very top of the income distribution.

Turning to wealth, it is well known that wealth is considerably more unequally distributed than income. The wealth Gini has fluctuated around 0.8 for most of the postwar period. It is also apparent that the Gini for wealth did not change much, if at all, between 1950 and 2007. By 2007, it stood at 0.82 and was only marginally higher than in both 1950 and 1971. However, the Gini coefficient increased substantially between 2007 and 2016.

Figure 5 shows the Gini coefficients together with 90% confidence intervals.[20] The Gini coefficients are tightly estimated, although the confidence bands are somewhat wider in the historical data. The observed long-run trends are clearly statistically significant. America is considerably more unequal today than it was in the 1970s, with respect to both income and wealth.

Figure 5: Gini coefficients with confidence bands



(a) Income (b) Wealth

Notes: Gini coefficient of income (panel (a)) and wealth (panel (b)) with 90% confidence bands. Confidence bands are shown as gray areas, and point estimates are connected by lines. Confidence bands are bootstrapped using 999 different replicate weights constructed from a geographically stratified sample of the final dataset.

[20] All confidence bands are computed using 999 replicate sample weights. Replicate weights are provided for the modern SCF surveys after 1983. For the historical surveys, we construct comparable 999 replicate weights. We compute sample weights for each draw of a geographically stratified sample from the final data after imputations and adjustments.

3.2 Income and wealth shares

We start the exploration of changes in income and wealth shares at the top, following the recent literature. The HSCF data corroborate the trajectories of U.S. income and wealth distribution that emerged from the well-known studies by Piketty and Saez (2003) and Saez and Zucman (2016). In a second step, we will use the more granular HSCF data to provide new evidence for distributional trends within the bottom 90% of the population.

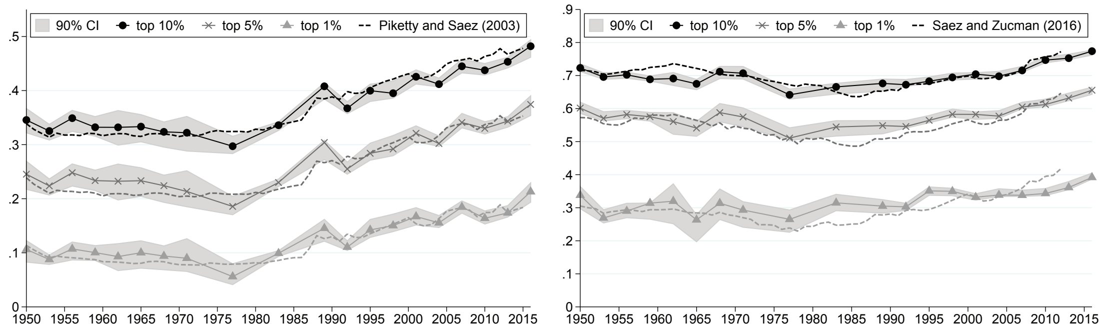
Figure 6a compares the income shares of the top 10%, 5%, and 1% of the income distribution in the HSCF to those calculated by Piketty and Saez (2003) using IRS data.[21] On the right-hand side, Figure 6b compares top wealth shares from the HSCF with those from Saez and Zucman (2016). Despite some minor discrepancies, it is clear that both the tax data and the HSCF data tell a similar story about the long-run trajectory of wealth and income inequality in postwar America. The increase in wealth inequality since the 1990s initially appeared somewhat stronger in the capitalized income tax data, but the gap has narrowed substantially with the 2016 SCF data. Kopczuk (2015) provides a detailed discussion of this phenomenon. We also note some small differences in the trajectory of the wealth distribution in the earlier decades between the IRS and the HSCF data. One reason for the divergence could be that Saez and Zucman (2016) had to adjust the pre-1962 estimates as households have been sorted by income rather than wealth. In Figure B.2 of the appendix, we consider income concentration among wealth-rich households and wealth concentration among income-rich households that point in this direction. Yet overall, administrative and survey data paint a similar picture of a marked increase in income inequality since the mid-1970s, and an increase in wealth inequality that is concentrated in the last decade.

A potential concern could be that the historical data provide too few or too noisy observations to allow for reliable inference at the top of the income and wealth distribution. We think that such concerns are likely unfounded. Figure 6 also shows estimated 90% confidence bands resulting from sampling error in the HSCF data for the top income and wealth shares. The top 10% income and wealth shares are tightly estimated. We also report the confidence intervals for the top 5% and top 1%, although these groups are not at the focus of our analysis. The confidence bands underscore that the reported increases in income and wealth inequality are statistically significant.

In the next step, we turn to the evolution of income and wealth across the entire distribution. The mirror image of increasing concentration of income in the hands of the top 10% must, by definition, be (relative) income losses among the bottom 90%. But which strata of the bottom 90% were hit particularly hard by the growing income share of the top 10%?

[21] Piketty and Saez (2003) include salaries and wages, small business and farm income, partnership and fiduciary income, dividends, interest, rents, royalties, and other small items reported as other income.

Figure 6: Top income and wealth shares



(a) Income (b) Wealth

Notes: Top income and wealth shares from HSCF data and Piketty and Saez (2003) and Saez and Zucman (2016). The dots show income and wealth shares from HSCF data, the dashed lines income shares from Piketty and Saez (2003) using IRS tax data or wealth shares from Saez and Zucman (2016) using IRS data and the capitalization method. The black dots show income (wealth) shares of the top 10%, dark gray crosses show the top-5% shares, and the light gray triangles show top-1% shares. Gray areas around the time series show 90% confidence bands. Confidence bands are bootstrapped using 999 different replicate weights constructed from geographically-stratified sample of the final data set. Horizontal axes show calender time and vertical axes income and wealth shares.

Table 4 shows the evolution of income and wealth shares of different parts of the population since World War II. As before, we report the income shares of different groups of the income distribution and wealth shares of different strata of the wealth distribution.[22] Starting with income on the left, the HSCF shows that the top 10% have grown their income share by close to 15 percentage points from 34.5% to 48.2% between 1950 and 2016. The 1970s and 1980s witnessed the strongest rise in the income share of the top 10% (+7.9 percentage points), mainly at the expense of the bottom 50%. In the 1990s and 2000s, the top 10% continued to expand their income shares (+4.1 percentage points), while above median income households (50%-90%) began to lose out more strongly while the income share of the bottom half stabilized. Overall, the income share of the bottom 50% of Americans has fallen by roughly a third from 21.6% to 14.5%, and middle-class households (50th to 90th percentiles) have lost about 6 percentage points in income shares. In other words, we do observe a hollowing out of middle-class America, with households around the median having witnessed the largest relative income losses.

The right side of Table 4 studies the change in wealth shares (households are now stratified by wealth). The main insight here is that until the 2008 financial crisis, changes in wealth shares were modest. If anything, the bottom 90% wealth share was slightly higher in 2007 than it was in 1950, and very close to its 1971 level. In contrast to the observed changes in

[22]Online appendix II reports the full time series.

Table 4: Shares in aggregate income and wealth

	Income					Wealth				
	1950	1971	1989	2007	2016	1950	1971	1989	2007	2016
bottom 50%	21.6	21.6	16.2	15.4	14.5	3.0	3.0	2.9	2.5	1.2
0%- 25%	6.1	6.2	5.0	4.5	4.5	-0.1	-0.2	-0.1	-0.1	-0.4
25%-50%	15.5	15.4	11.3	11.0	10.1	3.1	3.2	3.0	2.6	1.6
50%-90%	43.9	47.7	43.8	40.3	37.9	24.7	26.3	29.5	26.0	21.5
50%-75%	23.5	24.9	22.5	20.3	18.4	9.8	10.5	11.7	10.2	7.2
75%-90%	20.4	22.8	21.4	20.0	19.5	14.8	15.8	17.8	15.8	14.3
top 10%	34.5	30.7	39.9	44.3	47.6	72.3	70.7	67.6	71.5	77.4

the income distribution, middle-class households managed to maintain their wealth shares until the mid-2000s. The 50%-90% wealth share was higher in 2007 than in 1950, and only slightly lower than in 1989. It is equally clear that the financial crisis had a substantial effect on the wealth distribution. Middle-class wealth shares collapsed across the board, while the wealth share of the top 10% surged by 6 percentage points within less than a decade. In the HSCF data, the decade since the financial crises witnessed the largest spike in wealth concentration in postwar America.

To sum up, Table 4 shows that income concentration at the top rose strongly from 1970 to 1990 and has kept increasing since. By contrast, the wealth share of the bottom 90% rose until 1990 and fell markedly only after the 2008 financial crisis. The overall outcome was a more pronounced shift in the income distribution than in the wealth distribution since the 1970s. We return to this important fact in section 4.

3.3 Demographic change

What were the effects of secular changes in terms of educational attainment, age structure, and household size of the U.S. population on income and wealth inequality? Using the demographic information in the HSCF, we can disentangle these effects. We implement an approach proposed by Fortin, Lemieux, and Firpo (2011) to remove changes in the age structure and educational attainment over time and account for changes in household size by adjusting income and wealth at the household level to per-adult equivalents using the OECD equivalent scales.

The method we use relies on pooling data and calculating the probability of being surveyed in a base year (here, 1971) with the help of a probit regression. This method makes it possible to compute counterfactual inequality measures fixing demographic characteristics to the base year. As explanatory variables, we include age, educational attainment, the number of adults and children in a household, and the race of the household head. We use the estimated probability to reweight observations in other survey years to keep the demographic structure constant by multiplying the survey weights with the estimated probability.[23]

We first fix educational attainment and the age structure. Figure 7 shows Gini coefficients for the original data and the two counterfactual cases. The age effects on income are small, but the effects of education are quite sizable. This finding is in line with a rising college wage premium and an increase in the number of college-educated households. By contrast, the experience premium and life-cycle income profiles changed too little to have a notable impact on the Ginis. In the case of wealth, the effect of changing educational attainment is rather small, but the effect of aging is more pronounced. All in all, demographic changes have some effects but do not change the overall pattern of income and wealth inequality in the United States since World War II.

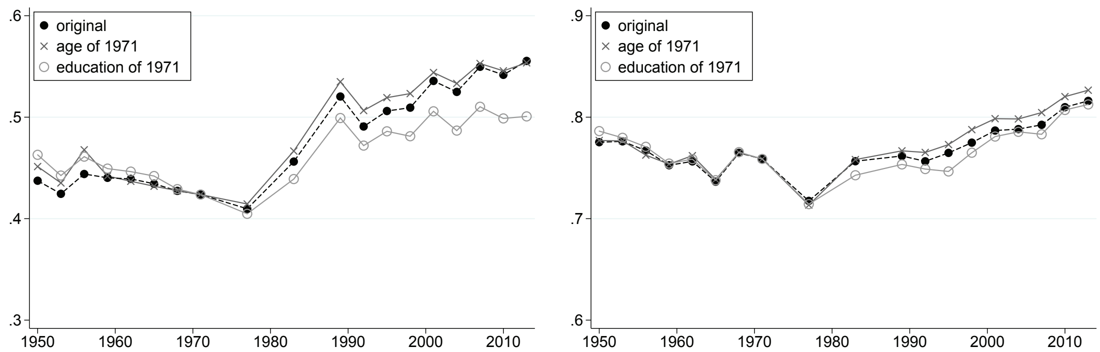
A second secular trend in the United States has been the decrease in average household size from an average of 3.4 household members in 1949 to an average of 2.5 in 2013 according to Census data. Given that the HSCF is a household survey, changes in household size can potentially affect measures of household-level inequality. To see if this is the case, we adjust income and wealth to per-adult-equivalent member of the household with the OECD equivalence scale.[24] Figure 8 reports Gini coefficients with and without adjustment for household size. Income concentration at the top falls somewhat when we look at adult-

[23]Reweighting factors are calculated in the following way: $D_Y = 0$ is a dummy indicating to which survey year the observation belongs. It is equal to 0 for the reference year and 1 otherwise. X are the explanatory variables. $\hat{P}(D_Y = 1|X)$ is the estimated probability of being surveyed in year Y given explanatory variables X. $\hat{P}(D_Y = 0|X)$ is the corresponding probability of being interviewed in the reference year. $\hat{P}(D_Y = 1)$ and $\hat{P}(D_Y = 0)$ are the sample proportions of households in the survey and reference year, respectively. The reweighting factor $\hat{\Psi}(X)$ is then given by

$$\hat{\Psi}(X) = \frac{\hat{P}(D_Y = 1|X)/\hat{P}(D_Y = 1)}{\hat{P}(D_Y = 0|X)/\hat{P}(D_Y = 0)}$$

[24]The OECD equivalence scale assigns a value of 1 to the first household member, 0.7 to each additional adult, and 0.5 to each child (see OECD `http://www.oecd.org/eco/growth/OECD-Note-EquivalenceScales.pdf`).

Figure 7: Gini coefficients accounting for demographic change



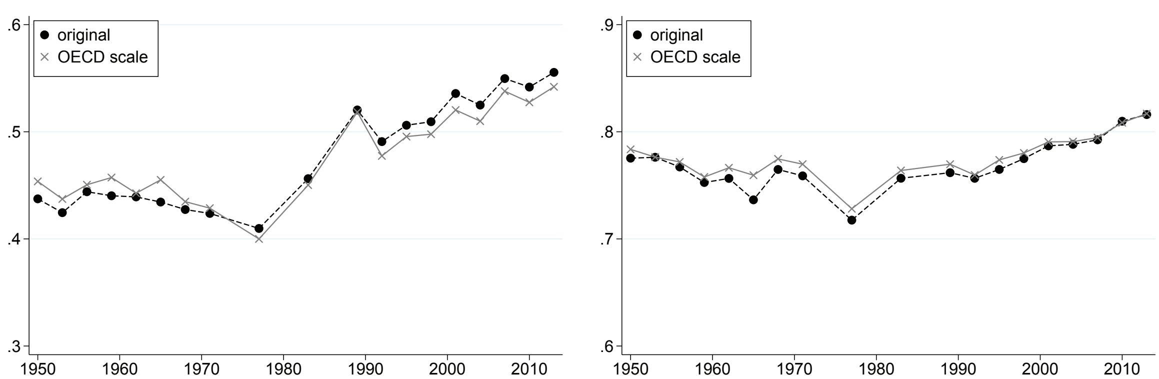
(a) Income Gini

(b) Wealth Gini

Notes: The two graphs show the effects of demographic changes on Gini coefficients. The black dashed lines are the results using the original data. For the dark gray solid lines with crosses, the age distribution is held constant at the 1971 distribution. For the light gray solid lines with dots, the distribution of education is held constant at the 1971 distribution. Age and education refer to head of household.

equivalent income. This trend is consistent with stronger assortative mating and increasing female labor force participation. For wealth, we do not observe big effects.

Figure 8: Gini coefficients for adult-equivalent income and wealth



(a) Income Gini

(b) Wealth Gini

Notes: The two graphs show Gini coefficients for adult-equivalent income and wealth. The black dashed lines are the results using the original data. For the gray solid lines with crosses, the data were adjusted with the OECD equivalence scale (see footnote 24).

3.4 The persistence of racial disparities in income and wealth

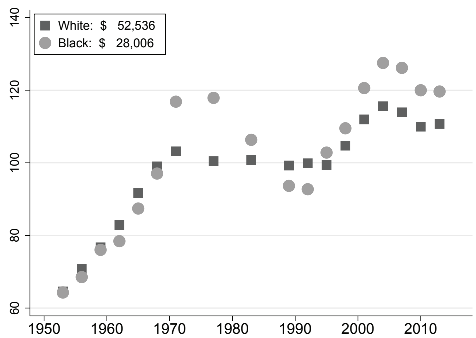
Race is an important stratifying dimension of the U.S. population. In a recent paper, Bayer and Charles (2017) provide new long-run evidence on the black-white earnings gap using data from the U.S. Census Bureau and the American Community Survey. They document persistent earnings differences for working-age men. The HSCF data provide a new perspective on the long-run evolution of racial inequality, complementing recent work along three dimensions. First, in addition to wage earnings, we can study household income from all sources. Second, our unit of observation is the household, not working-age male individuals. We thus capture the effects of changing marriage patterns, higher labor force participation of women, and rising incarceration rates of black men, as well as changes in transfers, education, and retirement decisions of households. Third, the HSCF data allow us to analyze the long-run evolution of wealth differentials between black and white households. So far, the racial wealth gap has remained uncharted territory as long-run data were simply not available. With data reaching back to the pre-civil rights era, our analysis extends recent work by Wolff (2017), who studied wealth differences between black, white, and Hispanic households in the modern SCF data starting in 1983.[25] For the analysis, we group households into black and white households according to the race of the household head. The number of interracial marriages is growing but remains small.[26] Following Bayer and Charles (2017), we not only compare the evolution of mean differences but also study racial disparities across different strata of the income and wealth distributions. Time series are indexed to 100 at the mean of the 1980 to 1989 period.

Figure 9 shows the trends in income and wealth of the median household and of the household at the 90th percentile for both white and black households. In the figure, a reduction in the racial divide will show up as a stronger increase in black households' income or wealth over time. A lockstep evolution of the series for black and white households signals persistence of racial disparities. Two facts stand out. First, income has grown at a comparable rate for black and white households. This means that pre-civil rights era disparities have largely persisted. Second, as the numbers indicate, the size of the racial income divide remains substantial. The median black household has about half of the income of the median white household. Third, the wealth gap is much larger than the income gap. The median black

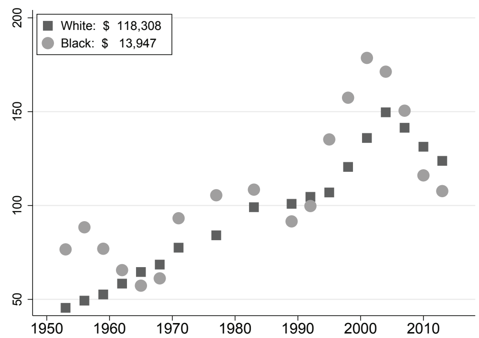
[25] Barsky, Bound, Charles, and Lupton (2002) use data from the Panel Study of Income Dynamics (PSID) and develop a nonparametric alternative to the Oaxaca-Blinder decomposition. They find that two thirds of the mean wealth racial gap can be accounted for by earnings differences. In their book, Oliver and Shapiro (2006) provide discussion of the historical sources of black-white wealth inequality with a detailed description of the historical institutional framework and discuss the available evidence on racial income and wealth inequality over time.

[26] For instance, Fryer (2007) reports that for whites, about 1% of marriages were interracial, for blacks, about 5%.

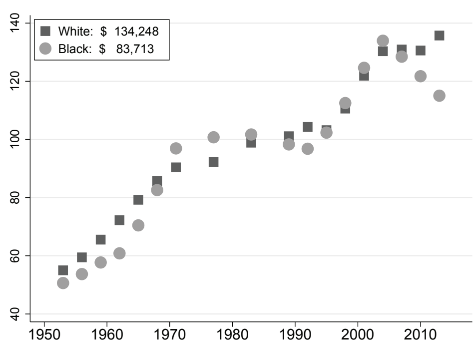
Figure 9: Black and white income and wealth trends



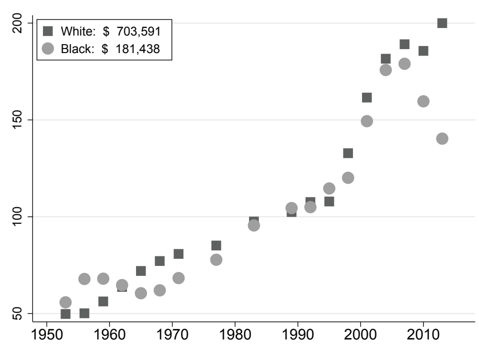
(a) Median income



(b) Median wealth



(c) Income 90th percentile



(d) Wealth 90th percentile

Notes: Panels (a) and (b) show median income and wealth of black and white households over time. Panels (c) and (d) show the 90th percentiles for income and wealth of black and white households over time. We show moving averages over three neighboring observations. Dark gray squares refer to white household heads and light gray dots to black household heads. Average wealth levels for 1983-1989 (in 2016 dollars) are shown in the legend.

household disposes of 12% of the wealth of a median white household. The wealth of the median black household stands at about $15,000 in 2016 prices — equivalent to the value of a car. The median white household has about $140,000 — corresponding to the value of a small house.

Looking at the time trends in more detail, we note two periods when the racial disparities narrowed temporarily. In the 1970s, the income of the median black household grew about 20% faster than the income of the median white household. However, the trend reversed in the 1980s when the share of black households headed by women increased strongly (and incarceration rates soared). In the 1980s, about 5 out of 10 black households were headed by a single female, up from 30% in the 1960s.[27] The 2000s are the second period in which

[27] When adjusting incomes for household size, the decline in relative incomes for black households during the 1980s becomes much less pronounced.

the racial income gap narrowed for the median household.

Figure 9b exposes an equally persistent racial wealth gap. The difference in wealth narrowed temporarily in the housing boom of the 1990s and early 2000s, but widened again after the financial crisis. At the top of the wealth distribution, the 90th percentiles evolved in lockstep for a long time. However, after 2007, the wealth levels of households at the 90th percentile of the black wealth distribution collapsed, while the 90th percentile of the white wealth distribution remained largely unaffected. We will see that the portfolio of the 90th percentile of the black wealth distribution is similar to the median white household, which also experienced substantial wealth losses after 2007 (see Figure 9b).

As an alternative to looking at racial differences, Bayer and Charles (2017) introduced the concept of a racial "rank gap" to study the evolution of earnings differences over time. Adapted for wealth, the rank gap is the percentage point difference between the rank of a given percentile in the black and white wealth distribution. For instance, a number of -30 for the median of the black wealth distribution means that the place of that household would be 30 percentage points lower on the white wealth distribution, that is, only at the 20th percentile.

Figure 10: Racial wealth rank gaps

(a) Rank gap median

(b) Rank gap 90th percentile

Notes: Racial wealth rank gaps at the median and 90th percentile. The racial rank gap is the difference in percentage points between the rank of the median (90th percentile) in the wealth distribution of black households and the position that this wealth level takes in the distribution of white households. A negative value indicates that the median (90th percentile) from the distribution of black households is below the median (90th percentile) in the wealth distribution of white households. Dashed lines show the long-run average of the racial wealth rank gaps.

Figure 10a shows the wealth rank gap at the median and the 90th percentile. For the median, the long-run average is close to -30, implying that the median black household is at the 20th percentile of the wealth distribution of white households. Put differently, the typical black household is poorer than 80% of white households. The rank gap fluctuates over time, tracking what we have seen for levels in Figure 9b, but is highly persistent over time. We also find a large and persistent rank wealth gap at the 90th percentile. The 90th percentile of the black wealth distribution corresponds roughly to the 60th percentile of the wealth distribution of white households. As both of these rank orderings appear highly persistent over time, our main conclusion is that virtually no progress has been made over the past 70 years in reducing the wealth inequality between black and white households.

4 Asset prices and the wealth distribution

In the previous section, we discussed changes in the income and wealth distributions separately, as in the existing literature. Yet it is precisely the link between the income and wealth distributions that plays a central role in theoretical models of wealth inequality. A central advantage of the HSCF is that it allows us to study the long-run evolution of the *joint* distributions of income and wealth. This topic is what we turn to now.

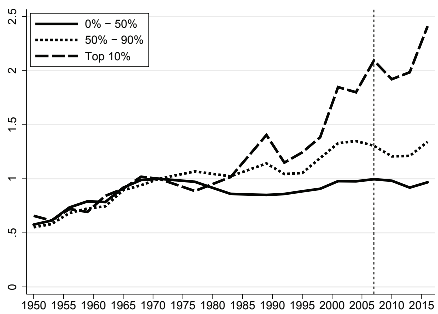
In the simplest model of the dynamics of the wealth distribution, changes in the income and wealth distributions are closely linked. With constant savings rates and identical returns on wealth, changes in the wealth distribution would be solely driven by changes in the income distribution. Or, put differently, the differential growth rates of wealth would be a function of the differential growth rates of income. Recent studies have questioned this assumption. Models based on labor income risk typically produce too little wealth concentration at the top and cannot account for substantial shifts in wealth inequality that occur over short time horizons (Benhabib and Bisin (2016), Gabaix, Lasry, Lions, and Moll (2016), Hubmer, Krusell, and Smith (2017)).

As a first check, in Figure 11 we compare the time path of income and wealth growth in the United States since 1971. Note that we stratify all households by wealth and index income and wealth levels to 1 in 1971. Figure 11a highlights a substantial divergence in income growth for different groups of the wealth distribution. Income growth was low for the bottom 90% and particularly meager for households in the lower half. For the bottom 50%, real incomes have stagnated since the 1970s. For households between the 50th to 90th percentiles of the wealth distribution, real incomes rose modestly by about a third over nearly 40 years, implying modest annual growth rates of much less than 1% per year. By contrast, income growth at the top was strong. The incomes of households within the top 10% of the wealth distribution have doubled between 1971 and 2007.[28]

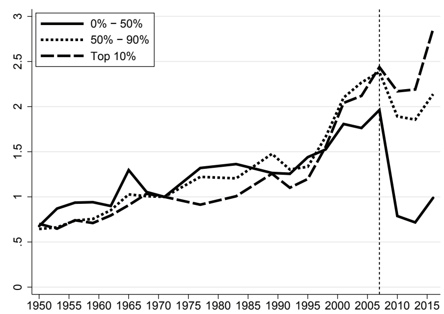
Yet when we turn to wealth growth for the same groups in Figure 11b, the contrast is stark.

[28] We report in appendix II the income shares for households along the wealth distribution comparable to the income shares along the income distribution in Table 4.

Figure 11: Income and wealth growth along the wealth distribution



(a) Income growth



(b) Wealth growth

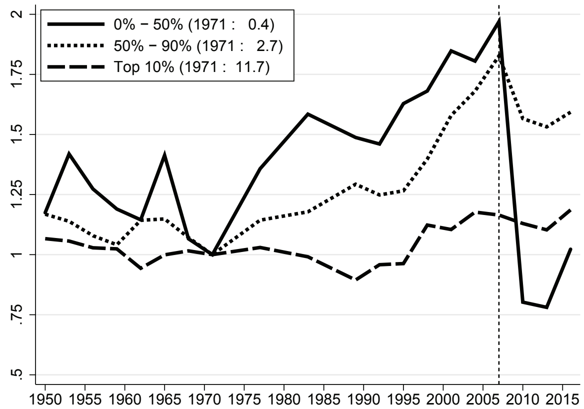
Notes: Growth of income and wealth for different wealth groups. All time series are indexed to 1 in 1971. The solid lines show growth rates for the bottom 50%, the short dashed lines for the middle class (50%-90%), and the long dashed lines for the top 10%. See text for further details.

From 1971 to 2007 (the last pre-crisis survey), wealth growth has been, by and large, identical for the top 10% and the bottom 90% of the wealth distribution. More precisely, middle-class (50%-90%) wealth increased by 140%, exactly at the same rate as the top 10% wealth. And even the bottom 50% did not do too badly when it comes to wealth growth, as their wealth still doubled between 1971 and 2007. Wealth and income growth rates have decoupled over an extended period, in marked contrast to the simple model sketched above. We will return to this point below.

Figure 11b also shows how devastating the 2007-2008 financial crash was to lower middle-class wealth. The impact of the crisis on wealth at the top was rather minor. By contrast, the wealth of the bottom 50% fell dramatically. By 2013, the absolute level of real wealth of the median household was 20% below its 1971 level. Within a few years, the crisis wiped out all gains in household wealth that the bottom 50% of the distribution had made over the preceding four decades. As of 2016, still close to half of the American population dispose of less wealth in real dollar amounts than in 1971.

One upshot is that before the financial crisis, wealth-to-income ratios increased most strongly in the middle and at the bottom of the wealth distribution and then also fell the most strongly for those groups during the crash. Figure 12 illustrates this phenomenon. The bottom 50% and the middle class experienced the strongest increase in wealth-to-income ratios, followed by a substantial decline. At the top, wealth-to-income ratios hardly changed.

Figure 12: Change in wealth-to-income ratios by wealth group, 1950-2016



Notes: Wealth-to-income ratios are constructed as a ratio of averages within each group. The solid lines refer to the bottom 50%, the short dashed lines to the middle class (50%-90%), and the long dashed lines to the top 10%. All series are indexed and show growth relative to 1971. Legend reports levels of wealth-to-income ratios for all wealth groups in 1971. See text for further details.

4.1 The dynamics of the wealth distribution

If we are to understand the dynamics of the wealth distribution in America over the past seven decades, we must look beyond income growth. In the following, we demonstrate that asset price changes played an important role in the observed dynamics of wealth inequality in postwar America.[29]

Asset prices affect the dynamics of the wealth distribution through two channels. First, asset prices lead to differential capital gains if portfolios differ across the distribution. We document this important stylized fact for the U.S. economy below. As changes in asset prices revalue existing wealth, they can induce shifts in wealth shares that are unrelated to income changes. Moreover, they can do so over short horizons as they immediately affect the value of accumulated assets. We will show that in America, persistent differences in portfolio composition between middle-class households and rich households essentially give rise to a race between the stock market and the housing market in shaping the dynamics of the wealth distribution.

The second channel through which asset prices matter for the dynamics of wealth inequality is through their effect on wealth-to-income ratios. The level of the wealth-to-income ratio determines the relative importance of savings flows for wealth dynamics. When wealth-to-income ratios are high, income growth and savings flows become relatively less important for

[29]For the French case, Garbinti, Goupille-Lebret, and Piketty (2017) show that price effects played an important role in shaping the French wealth distribution over the past 200 years. In the American context, Saez and Zucman (2016) discuss that price effects can change inequality trends relative to those implied by income and saving rate differences but focus on saving rate differences in their discussion.

the wealth distribution, simply because the stock of wealth is high relative to income flows. This second channel is particularly relevant for the time period from 1970 to 2007, when the aggregate wealth-to-income ratio increased from 4 to more than 7, driven by rising house prices and a booming stock market.

To understand how asset prices affect wealth inequality when portfolios are heterogeneous, consider a household i in period t with a portfolio of assets $\{A^i_{j,t}\}^J_{j=1}$, for instance, houses, stocks, and saving accounts. For household i, the capital gain Π^i_t from asset price changes between period t and $t+1$ is the asset-weighted average of price changes

$$\Pi^i_t = \sum_{j=1}^{J} \left(\frac{p_{j,t+1}}{p_{j,t}} - 1 \right) A^i_{j,t},$$

where $p_{j,t}$ denotes a (real) price index for asset j in period t. Denote the household's wealth in t by W^i_t and divide both sides of the equation by wealth to get

$$\begin{aligned} \frac{\Pi^i_t}{W^i_t} &= \sum_{j=1}^{J} \left(\frac{p_{j,t+1}}{p_{j,t}} - 1 \right) \frac{A^i_{j,t}}{W^i_t} \\ q^i_t &= \sum_{j=1}^{J} \left(\frac{p_{j,t+1}}{p_{j,t}} - 1 \right) \alpha^i_{j,t}, \end{aligned} \tag{1}$$

where $\alpha^i_{j,t}$ denotes the portfolio share $\frac{A^i_{j,t}}{W^i_t}$ of asset j for household i in period t and q^i_t is the growth rate of the household's wealth from capital gains. Equation (1) shows that portfolio differences (i.e., differences in $\alpha^i_{j,t}$ across households) lead to differences in capital gains q^i_t.

To fix ideas and structure the discussion about how this affects the wealth distribution, we rely on a simple but illuminating accounting framework adapted from Saez and Zucman (2016).[30] Consider a simplified law of motion for wealth of household i:

$$W^i_{t+1} = W^i_t(1 + r^i_t + q^i_t) + Y^i_t - C^i_t$$

where r^i_t are returns on wealth other than capital gains (e.g., dividends), Y^i_t denotes income from all other sources, and C^i_t denotes consumption. This approach simplifies wealth dynamics by abstracting from bequests and death, divorce, marriage or other life-cycle events that affect households' wealth accumulation.[31] The savings flow S^i_t of household i in period

[30] A microfounded analysis of household saving behavior and portfolio choice requires a more complex approach. Hubmer, Krusell, and Smith (2017) discuss why such a framework remains beyond reach for the time being but must become a topic for future research.

[31] We adjust the framework by Saez and Zucman (2016) slightly with respect to the timing convention by assuming that capital gains accrue together with savings flows, but the underlying mechanism remains the same. Saez and Zucman (2016) focus on the heterogeneity in savings behavior as they assume homogeneity

t corresponds to total income net of consumption $S_t^i = r_t^i W_t^i + Y_t^i - C_t^i$. Define further the saving rate s_t^i as $s_t^i = \frac{S_t^i}{Y_t^i}$, so that the law of motion for wealth becomes

$$W_{t+1}^i = W_t^i(1+q_t^i) + S_t^i = W_t^i(1+q_t^i) + s_t^i Y_t^i = (1 + q_t^i + \sigma_t^i)W_t^i \tag{2}$$

with $\sigma_t^i = \frac{s_t^i Y_t^i}{W_t^i}$. The growth rate of wealth comprises two components: the term σ_t^i capturing the contribution of savings to wealth growth and the term q_t^i capturing the effect of capital gains to wealth growth.

In the next step, we move from the law of motion for wealth *levels* to a law of motion for wealth *shares* of different wealth strata. We construct "synthetic" saving flows and capital gains for specific wealth groups, again taking the lead from Saez and Zucman (2016). Savings flows and capital gains for wealth groups are "synthetic" in the sense that they assume that households stay in their wealth group from one period to the next. While there is some mobility in practice, the synthetic method will yield a good approximation of wealth dynamics over the short run if households entering and leaving individual wealth strata are sufficiently similar. Demographic change and secular trends in saving behavior over the life cycle will require a more complex approach.

Note that we will now use i to refer to the group of households in a specific wealth stratum. The wealth share of group i in period t is $\omega_t^i = \frac{W_t^i}{W_t}$ where W_t is aggregate wealth in period t. All aggregate variables are defined according to group variables; for example, the aggregate savings rate is $s_t = \frac{S_t}{Y_t}$. Applying some straightforward transformations to equation (2) yields the law of motion for the wealth share ω_t^i:

$$\omega_{t+1}^i = \frac{1 + q_t^i + \sigma_t^i}{1 + q_t + \sigma_t}\omega_t^i \qquad \Longleftrightarrow \qquad \frac{\omega_{t+1}^i}{\omega_t^i} = \frac{1 + q_t^i + \sigma_t^i}{1 + q_t + \sigma_t}. \tag{3}$$

The law of motion has an intuitive interpretation: the wealth share of any group i increases if the group's wealth growth rate exceeds the average wealth growth rate in the economy. Differences in growth rates can result from the two components of wealth growth in equation (2). First, group i's capital gains q_t^i can be higher (or lower) than the average capital gain q_t in the economy. Second, different rates of wealth growth can also result from the difference between group i's savings component σ_t^i and average savings σ_t. This is the channel through which differences in income growth translate into wealth inequality: higher incomes of group i will, all else equal, increase the group's saving flows and its wealth level relative to other groups in the economy.

The savings term σ_t^i in the equation also comprises the inverse of the wealth-to-income ratio.

---

of capital gains. Our discussion focuses on the comparison of the relative importance of savings and capital gains as drivers of wealth accumulation.

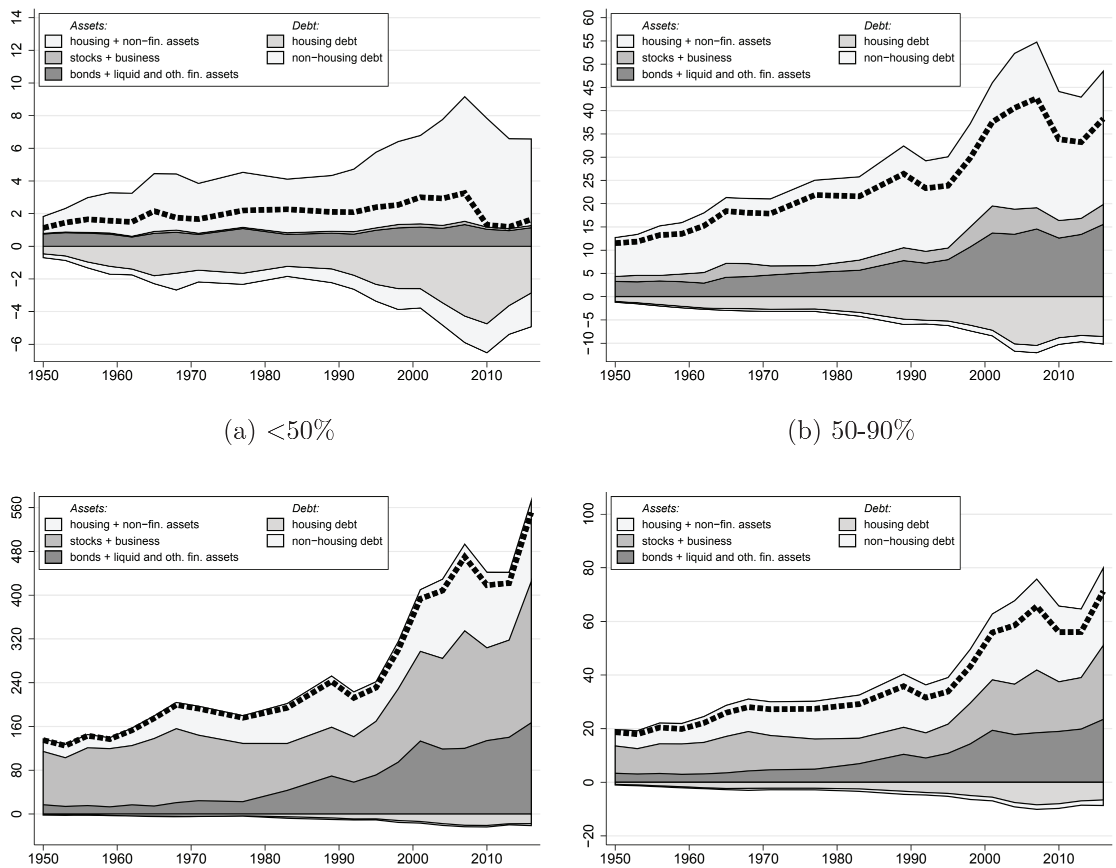
With higher wealth-to-income ratios, the importance of savings flows declines and tends to zero. In other words, the relative importance of differences in savings flows diminishes with higher wealth-to-income ratios. This effect is independent of portfolio composition and is operative even when households have identical wealth portfolios (so that capital gains are identical). Also note that the two components q_t^i and σ_t^i are independent, so that a low savings component relative to the average can go hand-in-hand with a large capital gains component relative to the average for group i, and vice versa. This can decouple the evolution of the income and wealth distributions.

4.2 Portfolio heterogeneity and asset price exposures

If portfolios differ systematically along the wealth distribution, asset price changes will lead to differential capital gains along the wealth distribution. These in turn can induce changes in the wealth distribution that are unrelated to changes in the income distribution. The necessary condition for such effects is that portfolios are heterogeneous. For the first time, the HSCF provides us with long-run balance sheet information to study the composition of household portfolios over a long time horizon. The evidence points to systematic and highly persistent differences in wealth portfolios across groups and hence a potentially important role for asset prices in shifting the wealth distribution, as we will now demonstrate.
Figure 13 displays the heterogeneity of household portfolios. It tracks the portfolio composition of the bottom 50%, the 50%-90%, and the top 10% of the wealth distribution since 1949. As a benchmark, we also track the average portfolio of the macroeconomy. In the figures, assets enter with positive values and debt as negative values. The household wealth of the wealth group corresponds to the consolidated value of all portfolio positions and is indicated by a dashed line in each of the figures. The degree of leverage in household portfolios can be inferred by looking at the sum of assets in excess of wealth. This systematic and persistent heterogeneity of household portfolios along the wealth distribution is an important, new stylized fact that the HSCF data reveal. We provide more detailed information in the online appendix IV.

Bottom 50%: The bottom 50% have very little wealth. Their small wealth position masks substantial gross positions. Houses and other nonfinancial assets, mainly cars, make up more than 80% of the asset side of the balance sheet. Financial assets play a minor role in bottom 50% portfolios. On the liability side, housing debt is the dominant form of debt, but compared to other wealth groups, the bottom 50% also have a high share of nonhousing debt. In recent years, education loans make up a growing share of this debt (Kuhn, Schularick, and Steins (2017)). Assets exceed wealth by a large margin, indicating a high degree of leverage

Figure 13: Heterogeneity of household portfolios



(a) <50%

(b) 50-90%

(c) 90-100%

(d) all households

Notes: Household portfolios for four wealth groups. Light gray areas show nonfinancial assets, dark gray areas financial assets, and negative areas housing and nonhousing debt. The dashed line indicates wealth. Panel (a) shows portfolio of the bottom 50% of the wealth distribution, panel (b) portfolio of the 50%-90%, and panel (c) portfolio of the top 10%. Panel (d) shows the portfolio of all households. Portfolio components are shown in 10,000 CPI-adjusted 2016 dollars. Wealth groups are separately defined for each survey year.

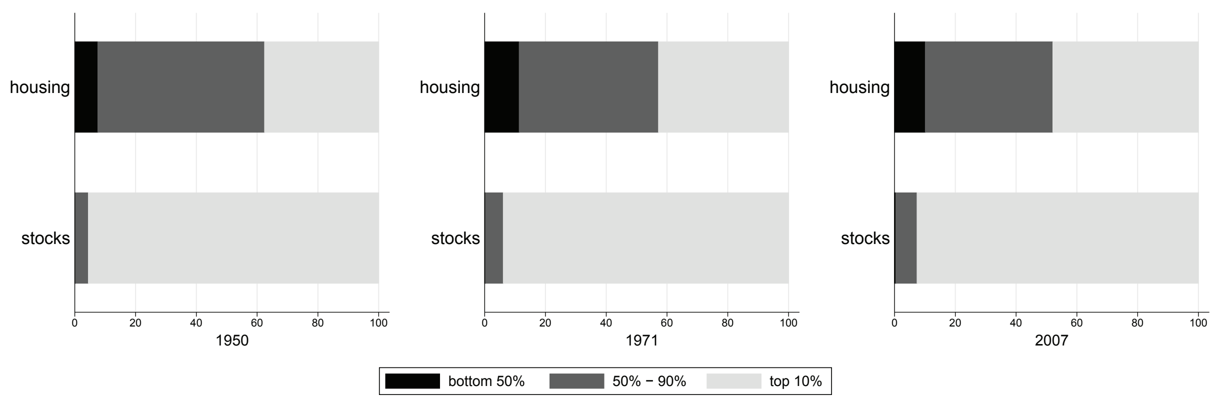
among the bottom 50%.

Middle class (50%-90%): The middle-class portfolio is dominated by nonfinancial assets. About two-thirds of the middle-class portfolio consists of houses and other nonfinancial assets. Direct stock holdings are typically below 5%. The large growth of other financial assets in the portfolio comes mainly from defined contribution pension plans. The middle class is also leveraged, with housing debt being the dominant debt component and assets exceeding wealth by 10% to 30%.

Top 10%: The top 10% are different when it comes to portfolio composition. The bulk of wealth is held in stocks and business equity. Houses as an asset class gained in importance for the top 10% but constitute a comparatively small fraction of assets. Other financial assets have grown strongly, mainly because of the proliferation of defined contribution pension plans. Leverage is low, so that for the top 10%, the sum of assets corresponds approximately to wealth.

Portfolio composition thus varies substantially and persistently along the wealth distribution. The portfolios of the bottom 90% are nondiversified and highly leveraged. Houses are *the* asset of the bottom 90%, making residential real estate the most egalitarian asset. Figure 14 highlights this point by showing the ownership structure of housing and stocks at different points in time.[32] The bottom 90% hold about half of all housing wealth but only a tiny fraction of stocks. Stocks are *the* asset of the wealthy in the sense that the top 10% consistently hold more than 90% of stocks. Looking at the Gini coefficient for individual asset classes confirms that housing is the most equally distributed asset and that the distribution of housing wealth has not changed substantially over time. We report Gini coefficients for the individual asset classes in Figure C.3 of the appendix.

Figure 14: Asset distribution by wealth group for selected years



Notes: Distribution of asset classes along the wealth distribution. Deposits include liquid assets and bonds, housing includes only the asset value of the house, and equity is stocks and business wealth. The black part of the bar on the left is the share of the bottom 50%, the gray bar is the share of households in the 50%-90%, and the light gray bar at the right is the share of the top 10%.

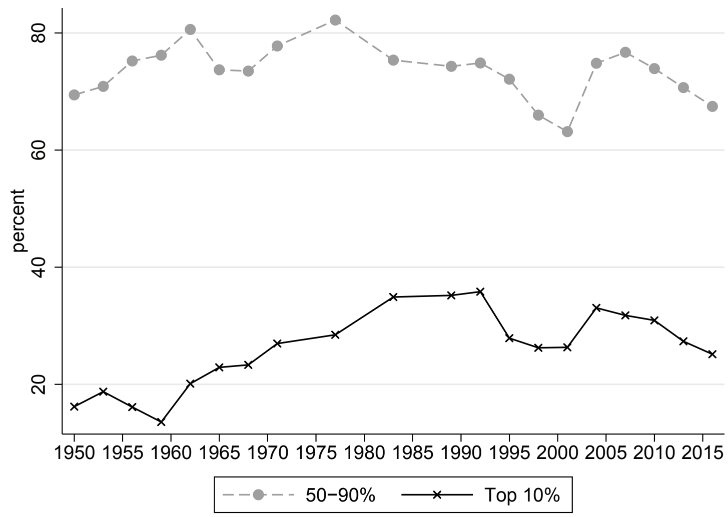
An important consequence of nondiversified and leveraged portfolio positions is that the wealth of middle-class households is highly exposed to changes in house prices. We quantify this exposure as the elasticity of wealth with respect to house prices, which is equal to $\frac{\text{Housing}}{\text{Wealth}}$, the ratio of the asset value of housing to wealth. Figure 15 shows the resulting exposure to house prices for middle-class households and households in the top 10% over time.[33] The

[32] We include mutual funds in the stock holdings. Results change little if we only consider direct stock holdings.

[33] Appendix D provides further results on house price exposure along the wealth distribution and changes

figure confirms that the elasticity of middle-class wealth to house prices is three to four times higher than at at the top. A 10% increase in house prices increases middle-class wealth by 6%-7%.[34] For changes in stock prices, the exposures are reversed. The top 10% are highly exposed, the rest very little.

Figure 15: House price exposure $\left(\frac{\text{Housing}}{\text{Net wealth}} \times 100\right)$ by wealth group



Notes: House price exposure for different wealth groups. House price exposure is measured by the elasticity of household wealth with respect to house price changes. See text for details. Horizontal axis shows calendar time and vertical axis house price exposure in percentage points.

4.3 The race between the stock market and the housing market

Such pronounced portfolio differences between households along the wealth distribution give rise to what we call the race between the stock market and the housing market: owing to their larger exposure, the middle class gains relatively more than top-wealth households when house prices rise. All else equal, rising house prices make the wealth distribution more

over time.

[34] We do not show the bottom 50% in this graph because of their large exposure. Figure D.4 in the appendix shows how their exposure compares to the other two groups. We also show there how the house price elasticity of wealth can be further broken down into a *diversification component* that is determined by the share of housing in assets and a *leverage component* measured by the debt-to-wealth ratio

$$\frac{\text{Housing}}{\text{Wealth}} = \underbrace{\frac{\text{Housing}}{\text{Assets}}}_{\text{diversification}} \times \left(1 + \underbrace{\frac{\text{Debt}}{\text{Wealth}}}_{\text{leverage}}\right).$$

equal, while stock market booms have the opposite effect: they primarily boost wealth at the top and lead to a more unequal distribution of wealth.

To explore how important this race between the stock market and the housing market has been for the wealth distribution in postwar America, we estimate the following regression relating changes in the top 10% wealth share over the three-year survey intervals to asset price movements:

$$\Delta \log(\omega_{t+1}^{top10}) = \beta_0 + \beta_h \Delta \log(p_{t+1}^h) + \beta_s \Delta \log(p_{t+1}^s) + \varepsilon_t,$$

where Δ is the first-difference operator, $\Delta x_{t+1} = x_{t+1} - x_t$, the superscript h denotes house prices and the superscript s stock prices.

Table 5 reports the estimated coefficients for the baseline regression in the first column. The signs of the estimated coefficients demonstrate how the race between the housing market and the stock market shaped wealth dynamics: rising house prices are associated with a falling top 10% wealth share. Rising stock prices boost the top 10% wealth share. Note that in the baseline specification, the error term comprises all other effects related to differences in savings or wealth-to-income ratios.

We control for these factors in additional regressions in columns (2)-(4) by adding the income share of the top 10% and changes in the ratio of income to wealth as regressors. As expected, the coefficients β_h and β_s become larger and significance rises. Clearly, while the effects are economically large, the sample is small and statistical significance varies.

The estimated regression coefficients have an intuitive interpretation as they correspond to the average elasticity of the top 10% wealth share with respect to asset prices.[35] From the law of motion for the wealth share, we can derive the elasticity of the wealth share of group i with respect to the price of asset j:

$$\frac{\partial \left(\frac{\omega_{t+1}^i}{\omega_t^i} \right)}{\partial \left(\frac{p_{t+1}^j}{p_t^j} \right)} = (1 + q_t + \sigma_t)^{-1} \left(\alpha_{t,j}^i - \alpha_{t,j} \frac{\omega_{t+1}^i}{\omega_t^i} \right).$$

Figure 16 shows the time series for the elasticity constructed from the portfolio shares in the HSCF data. The house price elasticity of the top 10% fluctuates around a mean of -0.17, very close to the point estimate of -0.16 in the wealth share regression (4) above. All else equal, a 10% increase in house prices will lower the top 10% wealth share by 1.6%. Assuming a top 10% wealth share of 75%, this corresponds to a decrease in the top 10% wealth share of 1.2 percentage points. A hypothetical 40% increase in real house prices would reduce the

[35] The portfolio share $\alpha_h = \frac{H}{W}$ in Figure 15 corresponds to the elasticity of the wealth level.

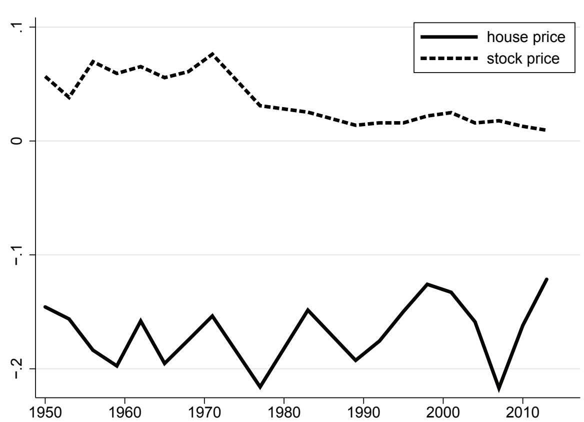
Table 5: The race between equity and house prices

	(1)	(2)	(3)	(4)
β_h	-0.104	-0.116	-0.138*	-0.157**
β_s	0.043*	0.044*	0.052**	0.043*
θ^{top10}	no	yes	no	yes
$\frac{Y}{W}$	no	no	yes	yes
N	19	19	19	19
R^2	0.162	0.246	0.352	0.468

Notes: Regression of changes in the top 10% wealth share on asset price growth and controls. Growth rates computed using log differences. θ^{top10} denotes the income share of the top 10% of households in the wealth distribution. $\frac{Y}{W}$ denotes controls for the inverse of the wealth-to-income ratio of the top 10% of households in the wealth distribution and for the aggregate economy. Asterisks show significance levels (* $p < 0.2$, ** $p < 0.1$, *** $p < 0.05$). All observations from the surveys from 1950 to 2016 are used.

wealth share of the top 10% from 75% to 70%, bringing it back to its 1971 level. For stock prices, the long-run average elasticity stands at 0.036 and is only slightly lower than the point estimate from the regression of 0.043. A 130% real increase in the stock market — comparable to the period between 1998 to 2007 — increases the wealth share of the top 10% by about 6 percentage points. This finding shows that asset prices have first-order effects on the wealth distribution.

Figure 16: Asset price elasticity of top 10% wealth share



Notes: Elasticity of the top 10% wealth share with respect to stock and house prices. See text for details.

4.4 Wealth gains from asset prices

The results of the previous section demonstrate that over short horizons, asset price fluctuations are closely associated with changes in wealth shares. We now turn to the cumulative effects that asset prices had on the wealth distribution over the past four decades. More precisely, we quantify the contribution of asset price changes to wealth accumulation of different groups.

We concentrate on wealth growth over two distinct periods. The first period comprises the nearly four decades from 1971 to the 2007-2008 financial crisis (the last pre-crisis survey was carried out in 2007). This was a period in which income distribution widened substantially, but measures of wealth inequality changed very little. We will see that house-price-induced wealth gains for the bottom 90% of the population played a central role in the observed stability of the wealth distribution.

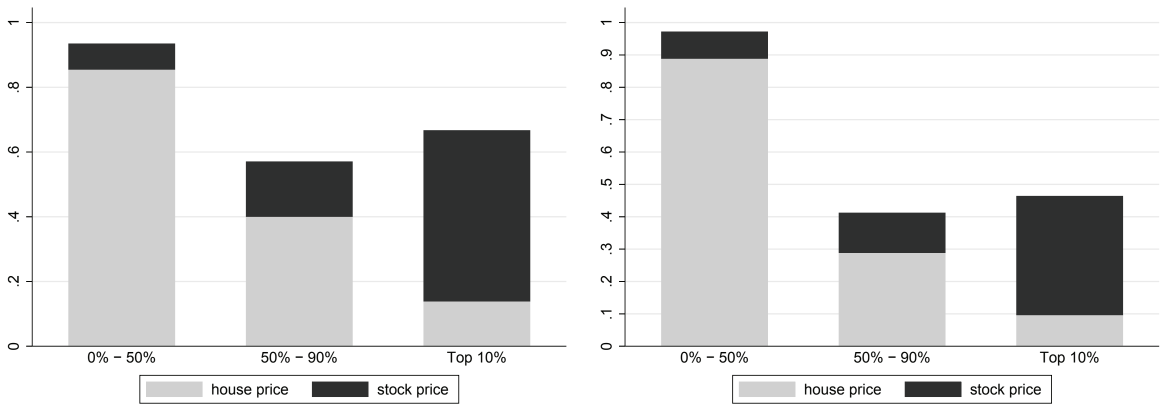
The second period that we study covers the decade after the financial crisis. As discussed above, this decade has witnessed the largest increase in wealth inequality in postwar history. The income distribution, by contrast, changed only modestly over this period. We will see that asset prices are again central in accounting for the large shift in the distribution of wealth over a relatively short period. Essentially, the quick recovery of the stock market boosted wealth at the top, while the middle class took much longer to recover from the substantial drop in wealth during the crisis.

Figure 17a shows how much wealth grew between 1971 and 2007 because of price changes in the stock market and the housing market. As before, we track wealth growth across three different wealth groups: the bottom 50%, the 50%-90%, and the top 10%. While Figure 17a displays the wealth growth from asset price changes, Figure 17b highlights the share of such asset-price-induced wealth growth in total wealth growth, including changes in wealth coming from savings flows.

Two observations stand out. First, over the entire period, wealth gains from rising asset prices were substantial across the distribution. As Figure 17a shows, the wealth of the bottom 50% grew by 90% only because of price effects. Also for the 50%-90% group and the top 10%, asset price changes induced sizable wealth growth of about 60%. It is easy to spot the race between the stock market and the housing market in the data: wealth growth from house price effects dominate for the bottom 90% of the wealth distribution, and stock price gains account for the bulk of the wealth gains at the top.

How much of the total wealth increase in the different groups is explained by price effects? Figure 17b shows that for the bottom 50% virtually all wealth growth over the 1971-2007 period came from higher asset prices. But even in the middle and at the top, asset prices still account for over 40% of total wealth growth, with the rest accounted for by savings. Note

Figure 17: Wealth growth from asset prices, 1971-2007



(a) Wealth growth from asset prices

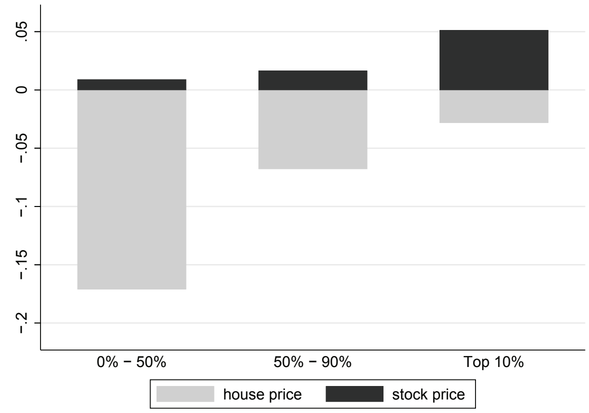
(b) Share in total wealth growth

Notes: Wealth growth component from the housing market and the stock market (q_t^i) in levels and as share of total growth for the bottom 50%, 50%-90%, and top 10% of the wealth distribution for the period from 1971 to 2007. The growth component in panel (a) is computed by fixing the housing and equity position at the beginning of the time period and then adjusting asset prices. Asset price gains or losses are expressed relative to the initial wealth level of the respective group. Panel (b) shows wealth growth component from asset prices as share of total wealth growth over the period from 1971 to 2007 for the different wealth groups.

that these estimates for the role of asset prices for wealth growth are likely conservative as households increased their exposure to the housing market over this period. In other words, wealth gains from asset prices quantitatively played a roughly equally important role for the evolution of the wealth distribution as savings. The variation in the contribution to wealth growth along the distribution helps us to understand why wealth and income growth decoupled in the decades before the crisis: relative to income, wealth grew most strongly for the bottom 90%.

Asset price movements also explain why wealth concentration spiked after the financial crisis. House prices plummeted after 2007 and recovered only slowly in recent years. By 2016, they were still 10% below their 2007 peak level. By contrast, the stock market recovered more quickly and then climbed to new peaks. By 2016, the main stock market indices were about 30% above their 2007 levels in real terms. Figure 18 shows the race between the housing market and the stock market between 2007 and 2016. The bottom 50% lost 15% of wealth relative to 2007 levels, mainly because of lower house prices. By contrast, the top 10% were the main beneficiary from the stock market boom and were relatively less affected by the drop in residential real estate prices. The consequence of substantial wealth losses at the bottom and in the middle of the distribution, coupled with wealth gains at the top, produced a large spike in wealth inequality.

Figure 18: Wealth growth from asset prices since 2007



Notes: Wealth growth component from the housing market and the stock market (q_t^i) for the bottom 50%, 50%-90%, and top 10% of the wealth distribution for the period 2007 to 2016. The growth component is computed by fixing the housing and equity position at the beginning of the time period and adjusting asset prices. Asset price gains or losses are expressed as share of the initial wealth level of the respective group.

How would the distribution of wealth in America look today without asset price effects? To construct a counterfactual, we use the law of motion for wealth levels from equation (2), keeping all parameters constant but adjusting the asset return term, q_t^i, so that nominal house prices (or stock prices) only increased with the rate of CPI inflation since 1971 (i.e., we keep prices constant in real terms). A few caveats are necessary as this is an accounting exercise, not a general equilibrium analysis. Our aim here is not to provide point estimates but to illustrate the potential of asset prices to shift the wealth distribution.

Table 6 shows the measured change in the wealth shares of the three groups relative to 1971 as well as the counterfactual change under two scenarios. In the first scenario, we keep real house prices constant. In the second scenario, we fix real stock prices at their 1971 level. The table shows counterfactual changes in wealth shares under these two corner assumptions.

The key message from Table 6 is that the asset price effects were potentially large. From 1971 to 2007, rising house prices slowed down wealth concentration in the hands of the top 10% by 3.1 percentage points relative to the counterfactual 3.9 percentage points increase. Put differently, without higher house prices, the middle-class wealth share would have been another 2.1 percentage points lower by 2007 and the increase in wealth concentration at the top four times higher than what we observe in the data. The house price crash after 2007 largely reversed these effects, but even in 2016, the observed increase in the top 10% wealth share of 6.7 percentage points was still about one-third lower than the counterfactual increase of 9.1 percentage points. in the absence of house price growth since 1971. Again, it is important to realize the magnitude of these shifts in wealth share. A difference of 2

Table 6: Changes in wealth shares relative to 1971

		1989	2007	2016
bottom 50 %	observed change	-0.1	-0.6	-1.9
	constant house prices	-0.3	-1.5	-2.6
	constant stock prices	-0.1	-0.2	-1.7
50% - 90%	observed change	3.2	-0.3	-4.8
	constant house prices	2.8	-2.4	-6.5
	constant stock prices	3.7	3.0	-1.3
Top 10%	observed change	-3.1	0.8	6.7
	constant house prices	-2.4	3.9	9.1
	constant stock prices	-3.7	-2.8	3.0

Notes: Changes in wealth shares relative to 1971 for different wealth groups. The first row for each wealth group shows the observed change in wealth shares (including changing house prices). The second row ("constant house prices") shows the change in wealth shares with constant real house prices. The third row ("constant stock prices") shows the change in wealth shares with constant real stock prices.

percentage points corresponds to over 14% of total annual household income.

The last row for each wealth group in Table 6 also reports the corresponding counterfactual for constant stock prices. Without the stock market boom, the top 10% wealth share would have been 3 percentage points lower in 2007 than in 1971, and even over the whole period, the middle class (50%-90%) would not have lost ground relative to the top. As mentioned above, these counterfactual simulations are suggestive only, but they highlight that the wealth distribution is highly sensitive to asset price dynamics. Valuation changes have the potential to induce major shifts in the wealth distribution that are unrelated to the evolution of the income distribution. In light of these first-order effects, they will have to play an important role in theoretical models of the dynamics of the wealth distribution.

5 Conclusions

This paper made three new contributions to the literature on income and wealth inequality. First, we introduced the *Historical Survey of Consumer Finances* (HSCF), a new household-level dataset covering the financial situation of U.S. households since World War II. The HSCF complements existing datasets for long-run inequality research that are based on income tax and social security records. It also opens up new and important avenues for research. In particular, the HSCF makes it possible to study the joint distributions of income and wealth over time as it contains both income and balance sheet information, coupled with extensive demographic information. We expect that the new data will become a valuable resource for future empirical and theoretical research on inequality, household finance, political economy, and beyond.

Second, we exploited the new data to study the trends in income and wealth inequality. Previous research documented a trend toward increasing polarization of income and wealth since the 1970s. The new data confirm this finding and underscore that the American middle class was the main loser of increasing income concentration at the top. We also track the racial wealth gap between black and white households over the long run. The picture that the HSCF paints is one of a persistent income and wealth divide between white and black households. In the mirror of the HSCF, next to no progress has been made in reducing differences in income and wealth, raising new questions as to the sources of this persistence.

The third main contribution of the paper is to expose systematic and highly persistent differences in portfolio composition and leverage of households along the wealth distribution. An important consequence of these differences is that asset price changes have first-order effects on the wealth distribution. They lead to capital gains and losses that induce shifts in the wealth distribution that are unrelated to changes in the income distribution. In particular, middle-class wealth in America is highly sensitive to fluctuations in real estate prices, so that rising house prices lead to wealth gains of middle-class households and decrease wealth inequality, all else equal. From the 1970s to the late 2000s, such housing wealth gains partly mitigated the sharp increase in income inequality. Despite stagnant income growth, middle-class wealth grew at a rate similar to wealth at the top.

The magnitude of changes in the wealth distribution induced by this asset price channel can be large. We estimate that until 2007, middle-class capital gains on residential real estate slowed down wealth concentration in the hands of the top 10% by about two-thirds. The housing bust of 2007-2008 was a watershed event as it hit the middle class particularly hard. By highlighting the crucial role that portfolio composition, leverage, and asset prices play for the wealth distribution, this paper opens up new avenues for future empirical and theoretical research on the determinants of wealth inequality.

References

Antoniewicz, R. L. (1996): *A comparison of the household sector from the flow of funds accounts and the survey of consumer finances*. Division of Research and Statistics, Division of Monetary Affairs, Federal Reserve Board.

Attanasio, O. P., and L. Pistaferri (2016): "Consumption Inequality," *Journal of Economic Perspectives*, 30(2), 3–28.

Bach, L., L. E. Calvet, and P. Sodini (2016): "Rich Pickings? Risk, Return, and Skill in the Portfolios of the Wealthy," Discussion Paper 11734, Centre for Economic Policy Research.

Barsky, R., J. Bound, K. K. Charles, and J. P. Lupton (2002): "Accounting for the Black-White Wealth Gap: A Nonparametric Approach," *Journal of the American Statistical Association*, 97(459), 663–673.

Bayer, P., and K. K. Charles (2017): "Divergent Paths: A New Perspective on Earnings Differences Between Black and White Men Since 1940," *The Quarterly Journal of Economics*.

Benhabib, J., and A. Bisin (2016): "Skewed Wealth Distributions: Theory and Empirics," Working Paper 21924, National Bureau of Economic Research.

Benhabib, J., A. Bisin, and M. Luo (2017): "Earnings Inequality and Other Determinants of Wealth Inequality," *American Economic Review*, 107(5), 593–97.

Bricker, J., A. Henriques, and P. Hansen (2018): "How much has wealth concentration grown in the United States? A re-examination of data from 2001-2013," Discussion paper, Discussion Paper.

Bricker, J., A. Henriques, J. Krimmel, and J. Sabelhaus (2016): "Measuring income and wealth at the top using administrative and survey data," *Brookings Papers on Economic Activity*, 2016(1), 261–331.

Burkhauser, R. V., S. Feng, and S. P. Jenkins (2009): "Using the P90/P10 index to measure US inequality trends with Current Population Survey data: A view from inside the Census Bureau vaults," *Review of Income and Wealth*, 55(1), 166–185.

Burkhauser, R. V., S. Feng, S. P. Jenkins, and J. Larrimore (2012): "Recent trends in top income shares in the United States: reconciling estimates from March CPS and IRS tax return data," *Review of Economics and Statistics*, 94(2), 371–388.

Castaneda, A., J. Díaz-Giménez, and J.-V. Ríos-Rull (2003): "Accounting for the U.S. Earnings and Wealth Inequality," *Journal of Political Economy*, 111(4), 818–857.

Fagereng, A., L. Guiso, D. Malacrino, and L. Pistaferri (2016): "Heterogeneity and Persistence in Returns to Wealth," Discussion paper, Discussion Paper.

Fella, G., and M. De Nardi (2017): "Saving and Wealth Inequality," *Review of Economic Dynamics*.

Fortin, N., T. Lemieux, and S. Firpo (2011): "Chapter 1 - Decomposition Methods in Economics," in *Handbook of Labor Economics*, ed. by O. Ashenfelter, and D. Card, vol. 4a, pp. 1 – 102. Elsevier.

Fryer, Roland G., J. (2007): "Guess Who's Been Coming to Dinner? Trends in Interracial Marriage over the 20th Century," *Journal of Economic Perspectives*, 21(2), 71–90.

Gabaix, X., J.-M. Lasry, P.-L. Lions, and B. Moll (2016): "The dynamics of inequality," *Econometrica*, 84(6), 2071–2111.

Garbinti, B., J. Goupille-Lebret, and T. Piketty (2017): "Accounting for wealth inequality dynamics: Methods, estimates and simulations for France (1800-2014)," Discussion paper, Discussion Paper.

Glover, A., J. Heathcote, D. Krueger, and J.-V. Rios-Rull (2017): "Intergenerational Redistribution in the Great Recession," .

Gottschalk, P., and S. Danziger (2005): "Inequality of wage rates, earnings and family income in the United States, 1975–2002," *Review of Income and Wealth*, 51(2), 231–254.

Heathcote, J., F. Perri, and G. L. Violante (2010): "Unequal we stand: An empirical analysis of economic inequality in the United States, 1967-2006," *Review of Economic Dynamics*, 13(1), 15–51.

Henriques, A., and J. W. Hsu (2013): "Analysis of wealth using micro and macro data: A comparison of the Survey of Consumer Finances and Flow of Funds Accounts," Discussion paper, FEDS Working Paper.

Herkenhoff, K. F. (2013): "The impact of consumer credit access on unemployment," *University of Califomia Los Angeles mimeo.*

Hubmer, J., P. Krusell, and A. A. Smith (2017): "The Historical Evolution of the Wealth Distribution: A Quantitative-Theoretic Investigation," Working paper.

Kaymak, B., and M. Poschke (2016): "The evolution of wealth inequality over half a century: The role of taxes, transfers and technology," *Journal of Monetary Economics*, 77, 1–25.

Kennickell, A. B., R. L. Woodburn, and D. A. McManus (1996): *Weighting design for the 1992 Survey of Consumer Finances*. Board of Governors of the Federal Reserve System.

Knoll, K., M. Schularick, and T. Steger (2017): "No price like home: global house prices, 1870–2012," *The American Economic Review*, 107(2), 331–353.

Kopczuk, W. (2015): "What Do We Know about the Evolution of Top Wealth Shares in the United States?," *Journal of Economic Perspectives*, 29(1), 47–66.

Kopczuk, W., E. Saez, and J. Song (2010): "Earnings inequality and mobility in the United States: evidence from social security data since 1937," *The Quarterly Journal of Economics*, 125(1), 91–128.

Kuhn, M., and J.-V. Ríos-Rull (2016): "2013 Update on the US Earnings, Income, and Wealth Distributional Facts: A View from Macroeconomics," *Federal Reserve Bank of Minneapolis Quarterly Review*, 37(1).

Kuhn, M., M. Schularick, and U. I. Steins (2017): "The Great American Debt Boom, 1949-2013," Discussion paper, University of Bonn.

Malmendier, U., and S. Nagel (2011): "Depression Babies: Do Macroeconomic Experiences affect Risk Taking?," *The Quarterly Journal of Economics*, 126, 373–416.

Oliver, M. L., and T. M. Shapiro (2006): *Black Wealth, White Wealth: A New Perspective on Racial Inequality*. Taylor & Francis, New York and London.

Piketty, T., and E. Saez (2003): "Income inequality in the United States, 1913–1998," *The Quarterly Journal of Economics*, 118(1), 1–41.

Piketty, T., E. Saez, and G. Zucman (2016): "Distributional National Accounts: Methods and Estimates for the United States," Discussion paper, National Bureau of Economic Research.

Saez, E., and G. Zucman (2016): "Wealth inequality in the United States since 1913: Evidence from capitalized income tax data," *The Quarterly Journal of Economics*, 131(2), 519–578.

Schenker, N., and J. M. Taylor (1996): "Partially parametric techniques for multiple imputation," *Computational Statistics & Data Analysis*, 22(4), 425–446.

Wolff, E. N. (2017): "Household Wealth Trends in the United States, 1962 to 2016: Has Middle Class Wealth Recovered?," Working Paper 24085, National Bureau of Economic Research.

A Data details

This section provides further details on the steps that have been taken to compile the *Historical Survey of Consumer Finances*. Section A.1 provides further discussion regarding the weight adjustment to account for nonresponse before 1983, section A.2 discusses the imputation in case variables are missing in a survey year of the historical data, and section A.3 provides the details on the construction of replicate weights to compute confidence intervals for estimates from the historical surveys. Section A.4 reports the number of household observations for the different survey years.

A.1 Weight adjustment to account for nonresponse

We describe in section 2.2 how we account for nonresponse at the top of the income and wealth distribution before 1983. As a proof of concept, we apply our adjustment to the 1983 data. In 1983, observations from the list sample can be identified in the data so that we can drop the list sample from the data. After dropping the list sample, we adjust the weights using our proposed adjustment method. Table A.1 compares results for income and wealth shares of the original sample including the list sample with those values obtained using our weight adjustment on the sample excluding the list sample.

Table A.1: Income and wealth shares of original and reweighted sample of SCF 1983

	income		wealth	
	original sample	reweighting	original sample	reweighting
top 10%-5%	10.8	12.2	12.1	15.5
top 5%-1%	13.2	12.6	22.8	24.7
top 1%-0.5%	3.0	2.1	7.4	6.2
top 0.5%-0.1%	4.5	1.9	11.4	6.2
top 0.1%	3.3	1.5	12.8	5.7

The results show that the adjustment works as intended. For income, it slightly overestimates shares between the top 10% and 5% and slightly underestimates the shares of the top 5% to 1%. The fit deteriorates toward the right tail above the top 1%. Deviations are, however, always less than 2 percentage points. For wealth shares, the picture is similar. After applying the weight adjustment, the shares up to the top 1% match reasonably well and the fit deteriorates within the top 1%. If anything, the resulting income and wealth concentration in the historical surveys will be understated with this method. This would further reinforce our key finding that income inequality rose strongly and wealth inequality hardly changed between 1971 and 2007. There has been a small decline in wealth concentration in the first three decades of our sampling period. However, we should also expect that at lower levels of income and wealth concentration, this adjustment will likely align even better with a sample that had oversampled households at the top of the distribution.

A.2 Imputation of missing variables

This section provides further details on the imputation of missing variables by predictive mean matching as described in Schenker and Taylor (1996). Following the modern SCF, we use multiple imputation and produce five imputed values for each missing variable. The imputation involves several steps. First, a linear regression model of the variable of interest is estimated on a sample with nonmissing observations. For each of the multiple imputations, a random realization of the regression coefficients is drawn using the estimated variance-covariance matrix. Using this coefficient vector and the linear regression model, a prediction for the variable of interest is generated. The predicted values on missing and nonmissing observations are compared to find nonmissing observations that produce the closest prediction. For each missing observation, we choose the three observations among the nonmissing observations that have predicted values most similar to the respective missing observation. Out of these three, we choose one observation randomly and assign the value of the variable of interest to the corresponding missing observation. Hence, the linear regression model is only used to define the distance between missing and nonmissing observations. The imputed values for the variables are all observed values. We refer to Schenker and Taylor (1996) for an in-depth discussion of the topic.

For each missing variable, several adjacent surveys could in principle be used as nonmissing samples for the imputation. In order to determine which adjacent survey years are most suitable for imputing missing values, we implement the following optimization before imputation. First, we determine all income, asset, debt, and demographic variables that are available in the year for which the variable is missing. For each combination of adjacent years, we then construct a subset of variables that are available both in the year with missing values and in the adjacent years. As the predictive accuracy decreased with the number of explanatory variables, we select those variables with the highest predictive power by using the lasso method. This method sets regression coefficients to zero for variables with small predictive power. For each combination of survey years, we then regress the variable of interest on those variables selected by the lasso method.[36] Finally, we calculate the R^2 for each regression. We use the R^2 as a measure of how well the combination of adjacent years is able to predict the missing variable. The combination with the highest R^2 is chosen for the

[36]Only survey years conducted less than 15 years before or after the missing year are considered. Out of these surveys, we choose the four closest to the missing year.

imputation. Tables A to E of the online appendix report the detailed combination of survey years and the adjacent survey years used in the imputation together with the R^2 from the regression.

A.3 Confidence intervals for estimates of top income and wealth shares

Administrative data from tax records have the advantage that they provide virtually complete coverage of the top of the income distribution, allowing for analysis of even very small groups at the top of the distribution. To derive wealth estimates using the capitalization method as in Saez and Zucman (2016) requires a model relating income flows to income stocks. The validity of the underlying assumptions cannot be tested for the United States. and the model might be sensitive to these assumptions. Bricker, Henriques, and Hansen (2018) provide an instructive discussion of this type of *modeling error* in the capitalization approach. They show that since the 2000s, modeling error is likely substantial for the capitalization method, even exceeding the variability of the survey estimate from the SCF. Survey data do not involve modeling error as income and wealth are observed directly from the answers of survey participants. The survey data, however, contain measurement and sampling error. To provide estimates of the variability that results from these errors, replicate weights for the modern SCF data have been construct and are provided to researchers. We follow this practice and constructed replicated weights for the historical surveys. When constructing replicate weights for the historical samples, we try to follow as closely as possible the construction of replicate weights for the modern samples. Because of computational feasibility, we provide replicate weights based on the imputed and weight-adjusted data. As a consequence, the steps from the imputation and weight adjustment do not factor into the estimated sampling variability. For these data, we draw 999 samples from the data stratified by geographical units and adjust weights to get a nationally representative sample. We use these replicate weights to construct all confidence bounds of our estimates shown in the main paper.

A.4 Number of household observations across survey years

Table A.2 reports the number of household observations for the different sample years. As described in section A.2, there are five implicates for each household observation in the final data. For the historical data, we pool sample years when surveys were conducted annually to increase the accuracy of our estimates. We show the number of observations for single survey years and after pooling sample years in Table A.2. The years highlighted in bold are used for all results in the main part of the paper using the pooled samples.

Table A.2: Number of household observations

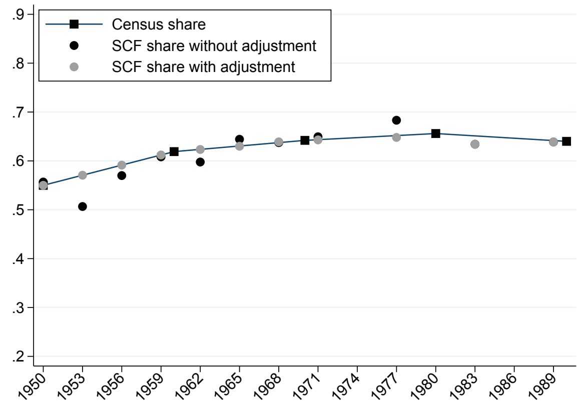
year	observations	pooled	year	observations	pooled	year	observations
1949	2,988		1961	1,799		**1983**	4,103
1950	2,940	5,928	**1962**	1,922	6,429	**1989**	3,143
1951	2,938		1963	1,819		**1992**	3,906
1952	2,435		1964	1,540		**1995**	4,299
1953	2,663	8,036	**1965**	1,349	4,708	**1998**	4,305
1954	2,599		1966	2,419		**2001**	4,442
1955	2,766		1967	3,165		**2004**	4,519
1956	2,660	8,025	**1968**	2,677	8,261	**2007**	4,417
1957	2,726		1969	2,485		**2010**	6,482
1958	2,764		1970	2,576		**2013**	6,015
1959	2,790	8,280	**1971**	1,327	6,388	**2016**	6,248
1960	2,708		**1977**	2,563		Total	110,497

Notes: Number of household observations in HSCF data. The first column shows the survey year. Survey years in bold are used for time series in the main paper. The second column shows the number of household observations for different survey years. The third column shows the number of observations after pooling survey years. For results in the main paper, pooled survey years are always used. Horizontal lines indicate the pooled annual survey years.

A.5 Homeownership rates

We explain in section 2.2 how we adjust survey weights to match population shares for age of the household head, college education, and race to be consistent with Census data. In the same step, we also target homeownership rates. Figure A.1 shows the homeownership rate in the Census (black squares) and in the HSCF with the adjustment of survey weights (gray dots) and without adjustment (black dots).

Figure A.1: Homeownership rates

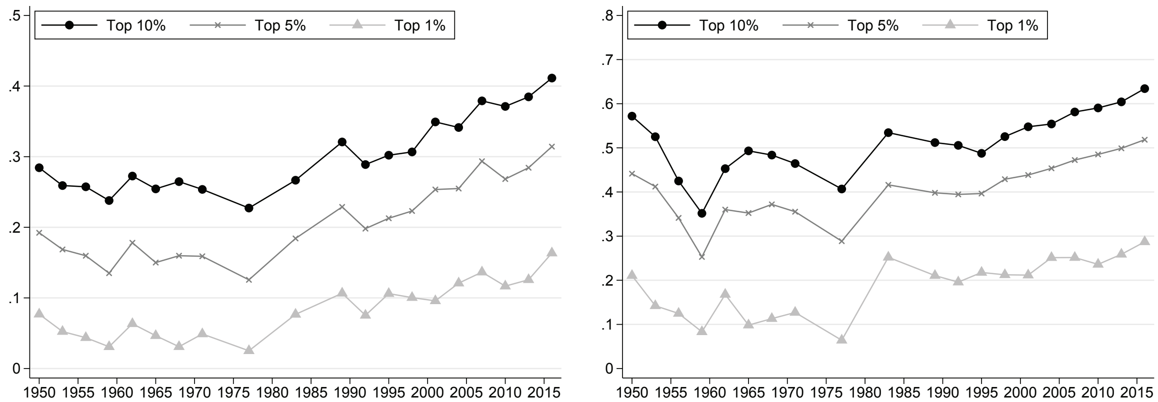


Notes: The large black squares show homeownership rates in the Census data. Census data are linearly interpolated in between years. The small black dots are the homeownership rates using the original survey data. The small gray dots are the homeownership rates using the adjusted survey data. The horizontal axis shows calendar time and the vertical axis population shares.

B Income concentration by wealth groups and wealth concentration by income groups

In the main paper, we report income and wealth concentration separately along the income and wealth distribution. There are important cases in which households that are at the top of the wealth distribution are not at the top of the income distribution and vice versa — for example, retired households that typically hold a lot of wealth but have little income. The HSCF data provide independent information on income and wealth, so we can explore the income concentration at the top of the wealth distribution and the wealth concentration at the top of the income distribution. Figure B.2, panel (a), shows the shares in total income of the top 10%, 5%, and 1% wealth-richest households over time. Panel (b) shows the shares in total wealth of the top 10%, 5%, and 1% income-richest households. Compared to Figure 6, the shares decline, yet the patterns with respect to the level of income and wealth concentration remain unaffected. Wealth is much more concentrated than income. Comparing the trends to those discussed before, the evolution of income and wealth concentration appears similar when we consider income concentration among the wealth-rich or wealth concentration among the income-rich.

Figure B.2: Shares in aggregate wealth and income



(a) Income Shares by Wealth Groups

(b) Wealth Shares by Income Groups

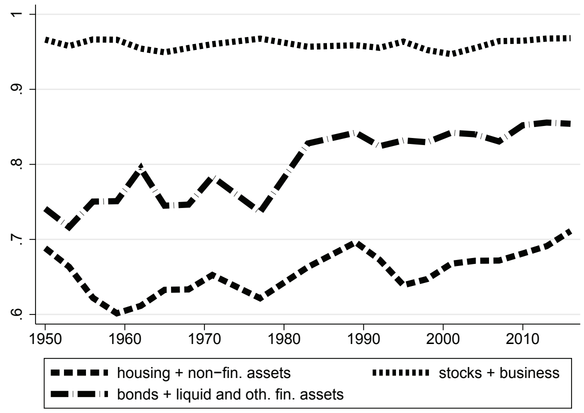
Notes: Panel (a) shows income shares for the top 10%, 5%, and 1% of the wealth distribution, panel (b) the wealth shares of the top 10%, 5%, and 1% of the income distribution.

C Gini coefficients for different asset classes

Section 4.2 in the main paper documents systematic differences in the household portfolios along the wealth distribution. We document that the distribution of stock holdings is highly skewed, while houses are in comparison relatively equally distributed. Here we offer an alternative view on the distribution of assets in the population by looking at Gini coefficients within asset classes. Kuhn and Ríos-Rull (2016) report for recent SCF data large differences in inequality of asset holdings within asset classes. The HSCF data allow us to extend such an analysis over the long run and document that such inequalities in the asset distributions have been a long-run phenomenon. Figure C.3 presents Gini coefficients for different asset classes. The underlying Gini coefficients are reported in Table H of the online appendix.

Corroborating the pattern from Figure 14, we find that housing is the most equally distributed asset, with a Gini coefficient varying between 0.6 and only recently exceeding 0.7. We observe a slight upward trend since 1960. By contrast, business equity and stocks show a very high degree of inequality, with Gini coefficients in excess of 0.95, and this pattern is also very stable over time. The Gini coefficients support the finding of polarized holdings of stock and business wealth and rather equally distributed housing assets.

Figure C.3: Gini coefficients for different asset classes over time



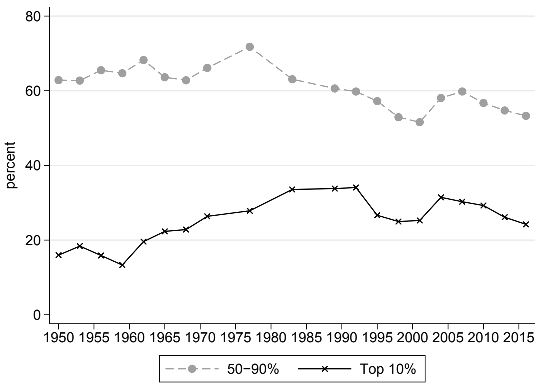
Notes: Gini coefficients for different asset classes over time. The solid line shows wealth, and the dashed line shows housing. The dashed-dotted line shows the sum of liquid assets, bonds, and other financial assets. The dotted line shows the sum of business wealth, other real estate, and equity. The horizontal axis shows the calendar time and the vertical axis the values of Gini coefficients.

D Decomposition of house price exposure

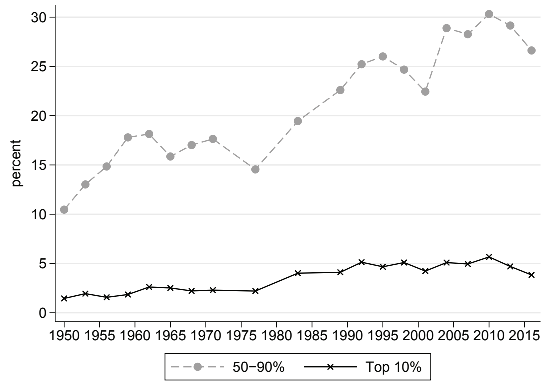
Figure D.4 shows the two components of house price exposure for the middle class and the top 10% over time. Panels (a) and (c) graph display the diversification component, panels (b) and (d) the leverage component. The share of housing in total assets of the middle class varies between 60% and 70% over time. For rich households, it varies between 10% and 20% and remains substantially lower than for the middle class throughout.

With respect to leverage, it is clear that the middle class is much more leveraged. Middle-class leverage increases from 20% in 1950 to a stunning 80% in 2010. Moreover, the strong exposure from low diversification and high leverage is not itself the result of rising house prices. Even in the 30 years between 1950 and 1980 — when real house prices were relatively stable (see Knoll, Schularick, and Steger (2017)) — the middle class held about 70% of its total assets in housing, and leverage amplified house price changes by approximately 40%.

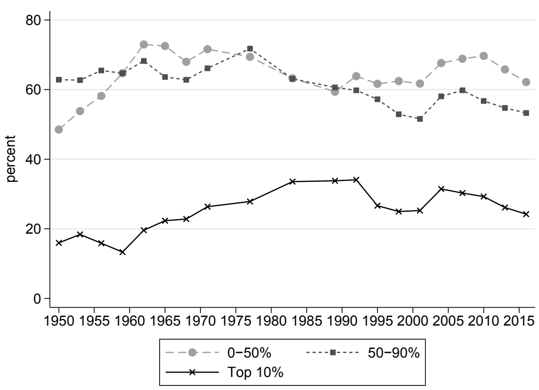
Figure D.4: Components of house price exposure by wealth group



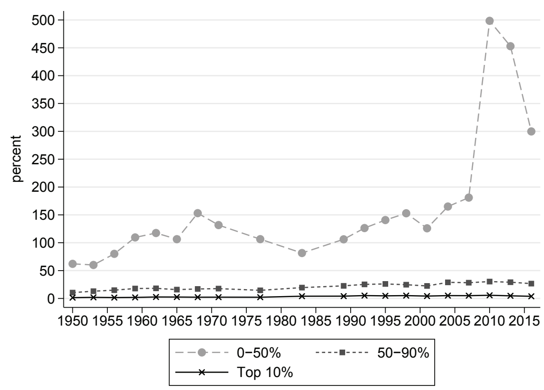
(a) Diversification ($\frac{\text{Housing}}{\text{Assets}} \times 100$)



(b) Leverage ($\frac{\text{Debt}}{\text{Wealth}} \times 100$)



(c) Diversification ($\frac{\text{Housing}}{\text{Assets}} \times 100$)



(d) Leverage ($\frac{\text{Debt}}{\text{Wealth}} \times 100$)

Notes: Decomposition of house price exposure for households in the bottom 50%, the 50%-90%, and the top 10% of the wealth distribution. Panels (a) and (c) show the *diversification* component, panels (b) and (d) the *leverage* component. See text for further details. Horizontal axes show calendar time, and vertical axes components are in percentage points.

Online Appendix
Not for Publication

This online appendix accompanies the paper "Wealth and Income Inequality in America, 1949-2016."

I Information on imputation of missing variables

Tables A to E provide information on the adjacent survey years used to impute missing variables in some of the survey years. We describe the imputation procedure in detail in Section A.2 of the appendix. In most cases, our imputation method selects a single survey year to impute missing information. This restriction to a single year is not predetermined as part of the imputation routine but rather the outcome that yields the best fit. We describe the method to select survey years as part of the imputation approach in the appendix.

Table A: Imputation of income variables

	survey year	years in imputation	R^2
labor income	1960	1959	97
	1961	1959	97
	1962	1959	96
	1963	1959	96
	1964	1966	88
	1965	1966	78
labor income + business	1971	1968	83
	1977	1968	84

Notes: The first column shows the name of the imputed variable, the second column shows the year for which imputation is done, and the third column shows the survey years from which information is used for the imputation. The number of years used for the imputation is not restricted to be one but chosen as part of the imputation routine. See description of imputation routine for further details.

Table B: Imputation of financial variables

	survey year	years in imputation	R^2
liquid assets	1964	1961	42
	1966	1968	38
bonds	1964	1963	42
	1966	1967	23
	1971	1970	67
equity	1948	1952	98
	1951	1952	73
	1954	1955	74
	1956	1955	75
	1957	1955	75
	1958	1962	76
	1959	1962	76
	1961	1962	77
	1965	1963	64
	1966	1968	52
	1971	1970	96

Notes: The first column shows the name of the imputed variable, the second column shows the year for which imputation is done, and the third column shows the survey years from which information is used for the imputation. The number of years used for the imputation is not restricted to be one but chosen as part of the imputation routine. See description of imputation routine for further details.

Table C: Imputation of cash value of life insurance

survey year	years in imputation	R^2
1948	SFCC1962	45
1949	SFCC1962	47
1950	SFCC1962	49
1951	SFCC1962	48
1952	SFCC1962	46
1953	SFCC1962	49
1954	SFCC1962	47
1955	SFCC1962	40
1956	SFCC1962	40
1957	SFCC1962	41
1958	SFCC1962	41
1959	SFCC1962	41
1960	SFCC1962	48
1961	SFCC1962	35
1963	SFCC1962	41
1964	SFCC1962	44
1965	SFCC1962	47
1966	SFCC1962	38
1967	SFCC1962	38
1968	SFCC1962	47
1969	SFCC1962	57
1970	SFCC1962	58
1971	SFCC1962	38
1977	SFCC1962	43

Notes: The first column shows the name of the imputed variable, the second column shows the year for which imputation is done, and the third column shows the survey years from which information is used for the imputation. The imputation is not restricted to use only SFCC 1962 data. Information on pension wealth is available in both the SFCC 1962 and the SCF 1983. The SFCC 1962 data are chosen as part of the imputation routine. See description of imputation routine for further details.

Table D: Imputation of nonfinancial variables

	survey year	years in imputation	R^2
value of home	1948	1951	42
	1952	1954	50
	1961	1960	30
other real estate	1948	1952	37
	1951	1952	59
	1954	1952	50
	1955	1952	57
	1956	1952	58
	1957	1962	50
	1958	1963	55
	1959	1963	55
	1961	1963	56
	1964	1963	61
	1965	1968	61
	1966	1963	50
	1967	1968	59
	1971	1968	54
business assets	1948	1953	48
	1949	1950	51
	1951	1953	52
	1954	1953	49
	1955	1953	50
	1956	1953	51
	1957	1953	51
	1958	1962	95
	1959	1962	94
	1961	1962	96
	1964	1962	96
	1965	1962	96
	1966	1970	30
	1967	1970	33
	1968	1963	61
	1969	1963	62
	1971	1962	94
	1977	1970	40

Notes: The first column shows the name of the imputed variable, the second column shows the year for which imputation is done, and the third column shows the survey years from which information is used for the imputation. The number of years used for the imputation is not restricted to be one but chosen as part of the imputation routine. See description of imputation routine for further details.

Table E: Imputation of debt variables

	survey year	years in imputation	R^2
housing	1948	1951	24
	1952	1954	45
	1961	1962	27
other real estate	1948	1949	72
	1952	1954	70
	1960	1959	88
	1961	1959	87
	1962	1959	87
	1963	1968	96
	1964	1968	88
	1965	1968	95
	1966	1968	81
	1967	1968	84
	1971	1968	94
nonhousing	1966	1968	29

Notes: The first column shows the name of the imputed variable, the second column shows the year for which imputation is done, and the third column shows the survey years from which information is used for the imputation. The number of years used for the imputation is not restricted to be one but chosen as part of the imputation routine. See description of imputation routine for further details.

II Time series on income and wealth shares

Table F shows income shares for three income and wealth groups over time. The groups are the bottom 50%, the 50%-90%, and the top 10%. Table 4 in the main paper shows the data for selected years. The last three columns show the income shares by wealth groups corresponding to the discussion in section 4 from the main paper.

Table F: Shares in aggregate income and wealth

	income shares by income groups			wealth shares by wealth groups			income shares by wealth groups		
year	0%-50%	50%-90%	top 10%	0%-50%	50%-90%	top 10%	0%-50%	50%-90%	top 10%
1950	21.6	43.9	34.5	3.0	24.7	72.3	33.1	38.4	28.4
1953	22.1	45.5	32.4	4.0	26.4	69.6	34.5	39.6	25.9
1956	20.7	44.4	34.9	3.8	26.0	70.2	34.9	39.4	25.7
1959	21.2	45.5	33.2	3.9	27.2	68.9	36.1	40.1	23.8
1962	21.1	45.7	33.2	3.4	27.5	69.1	33.8	38.9	27.3
1965	21.5	45.2	33.3	4.2	28.4	67.5	34.2	40.4	25.4
1968	21.6	46.1	32.3	3.1	25.7	71.1	34.2	39.4	26.5
1971	21.6	47.7	30.7	3.0	26.3	70.7	33.7	40.9	25.4
1977	21.5	48.8	29.7	4.0	31.9	64.1	33.1	44.1	22.7
1983	19.6	46.9	33.5	3.9	29.6	66.5	30.0	43.3	26.7
1989	16.2	43.8	39.9	2.9	29.5	67.6	25.8	42.1	32.1
1992	17.7	45.6	36.6	3.3	29.6	67.1	28.8	42.4	28.9
1995	16.4	45.3	38.3	3.5	28.2	68.2	28.6	41.2	30.2
1998	16.6	43.9	39.5	2.9	27.6	69.4	26.7	42.6	30.7
2001	15.9	41.7	42.4	2.7	26.9	70.4	24.6	40.5	34.9
2004	16.6	42.3	41.1	2.5	27.7	69.8	24.6	41.2	34.1
2007	15.4	40.3	44.3	2.5	26.0	71.5	24.0	38.1	37.9
2010	16.0	40.2	43.7	1.2	24.2	74.6	25.2	37.6	37.1
2013	15.3	40.1	44.6	1.1	23.7	75.3	23.7	37.8	38.5
2016	14.5	37.9	47.6	1.2	21.5	77.4	22.0	36.9	41.1

III Time series on Gini coefficients

Table G shows the time series of Gini coefficients over time. The table shows Gini coefficients every three years or between 1971 and 1983 for all available surveys. We discuss the observed time trends in section 3.1 of the main paper.

Table G: Gini coefficients for income and wealth

	Income			Wealth		
year	all	bottom 99%	bottom 90%	all	bottom 99%	bottom 90%
1950	44	39	31	81	74	61
1953	43	39	32	79	73	58
1956	46	41	34	79	73	58
1959	44	39	33	79	71	58
1962	44	40	33	80	72	61
1965	44	39	32	78	72	59
1968	43	39	32	81	74	62
1971	43	39	33	80	74	62
1977	42	40	34	76	70	60
1983	46	41	35	78	70	59
1989	53	46	39	79	72	63
1992	49	45	38	79	71	61
1995	52	46	39	79	70	60
1998	52	45	38	80	72	62
2001	54	46	38	81	74	63
2004	53	46	38	81	74	65
2007	55	47	38	82	74	63
2010	54	47	37	85	79	71
2013	56	48	38	85	79	71
2016	58	49	39	86	79	70

Notes: Gini coefficients for income and wealth for all households and bottom 99% and 90% of the income and wealth distribution. For the bottom 99% and 90% we exclude the top 1% and 10%, respectively, in the case of the income Gini of the income distribution, and in the case of wealth, from the wealth distribution.

Table H: Gini coefficients for different asset classes over time

year	housing + nonfin. assets	stocks + business	bonds + liq. and oth. fin. assets
1950	0.69	0.97	0.74
1953	0.66	0.96	0.72
1956	0.62	0.97	0.75
1959	0.60	0.97	0.75
1962	0.61	0.95	0.79
1965	0.63	0.95	0.74
1968	0.63	0.96	0.75
1971	0.65	0.96	0.78
1977	0.62	0.97	0.74
1983	0.66	0.96	0.83
1989	0.70	0.96	0.84
1992	0.67	0.96	0.82
1995	0.64	0.96	0.83
1998	0.65	0.95	0.83
2001	0.67	0.95	0.84
2004	0.67	0.96	0.84
2007	0.67	0.96	0.83
2010	0.68	0.96	0.85
2013	0.69	0.97	0.86
2016	0.71	0.97	0.85

IV Time series for portfolio composition for wealth groups

Tables I, J, and K show the portfolio composition of households for the four wealth groups considered in the main paper. These groups are the bottom 50%, the 50%-90%, and the top 10%. Portfolio shares are reported for surveys in three-year intervals. The first six columns show shares in assets, the next two columns show shares in debt, and the last column shows the debt-to-asset ratio.

Table I: Shares of wealth components in wealth portfolios of bottom 50% (in%)

year	other nonfin. assets	real estate	bus. wealth	equity	liquid assets, bonds	other fin. assets	non housing debt	housing debt	debt-to-asset ratio
1950	8.1	48.5	0.6	1.7	18.6	22.4	33.6	66.4	38.3
1953	8.6	53.8	0.2	1.5	17.1	18.7	31.7	68.3	37.5
1956	13.5	58.2	0.1	1.4	13.7	13.1	27.1	72.9	44.5
1959	10.6	64.7	0.1	2.0	12.7	9.9	28.2	71.8	52.3
1962	8.3	73.0	0.5	0.8	11.8	5.6	20.3	79.7	54.0
1965	7.2	72.5	0.2	2.3	8.9	8.8	21.2	78.8	51.5
1968	9.7	68.0	0.0	3.0	10.7	8.6	38.3	61.7	60.5
1971	7.8	71.6	0.1	1.8	10.0	8.7	32.9	67.1	56.8
1977	5.1	69.4	0.0	1.6	18.8	5.2	29.0	71.0	51.6
1983	16.8	63.2	1.0	1.5	10.3	7.2	33.2	66.8	44.9
1989	19.5	59.4	1.3	1.3	9.5	9.0	37.8	62.2	51.5
1992	17.3	63.9	2.0	1.1	7.6	8.1	32.6	67.4	55.8
1995	18.8	61.7	1.4	1.2	6.1	10.9	30.6	69.4	58.5
1998	16.8	62.5	1.4	1.8	6.3	11.2	33.0	67.0	60.5
2001	18.1	61.7	1.1	1.8	6.4	10.9	31.3	68.7	55.7
2004	16.0	67.6	1.0	1.3	5.2	8.9	28.3	71.7	62.3
2007	14.5	68.9	1.2	1.0	4.8	9.7	27.5	72.5	64.4
2010	15.3	69.7	1.3	0.4	4.3	8.9	27.2	72.8	83.3
2013	17.8	65.8	1.2	0.7	5.4	9.0	32.6	67.4	81.9
2016	18.6	62.2	1.2	0.8	6.0	11.2	42.0	58.0	75.0

Table J: Shares of wealth components in wealth portfolios of 50%-90% (in%)

year	other nonfin. assets	real estate	bus. wealth	equity	liquid assets, bonds	other fin. assets	non housing debt	housing debt	debt-to-asset ratio
1950	3.0	62.8	4.7	3.8	17.1	8.6	14.3	85.7	9.5
1953	3.1	62.7	4.8	5.5	15.0	8.8	14.7	85.3	11.5
1956	4.5	65.5	2.0	5.9	15.5	6.7	13.7	86.3	12.9
1959	4.8	64.7	1.3	9.1	14.7	5.4	13.7	86.3	15.1
1962	3.0	68.2	5.2	7.4	12.5	3.6	11.4	88.6	15.4
1965	2.8	63.6	3.5	10.5	13.8	5.7	12.3	87.7	13.7
1968	3.6	62.8	0.8	12.3	15.0	5.4	15.6	84.4	14.5
1971	2.5	66.1	1.7	7.6	15.6	6.5	14.3	85.7	15.0
1977	1.6	71.8	0.7	5.0	17.3	3.6	16.6	83.4	12.7
1983	6.4	63.1	5.8	2.6	13.3	8.7	19.4	80.6	16.3
1989	6.9	60.6	5.4	3.2	10.6	13.3	19.1	80.9	18.4
1992	6.9	59.8	5.0	3.7	10.2	14.4	13.9	86.1	20.1
1995	8.0	57.2	4.3	4.0	8.2	18.2	15.7	84.3	20.6
1998	6.8	52.9	4.6	6.9	8.9	19.8	16.6	83.4	19.8
2001	6.0	51.6	5.6	7.0	7.7	22.0	14.4	85.6	18.3
2004	6.0	58.1	5.1	5.2	7.0	18.7	13.4	86.6	22.4
2007	5.3	59.8	4.2	4.1	6.6	19.9	13.1	86.9	22.0
2010	6.2	56.7	5.1	3.4	7.2	21.3	13.9	86.1	23.3
2013	6.1	54.7	3.9	4.1	6.8	24.3	13.8	86.2	22.6
2016	5.8	53.3	4.5	4.3	7.1	24.9	16.1	83.9	21.0

Table K: Shares of wealth components in wealth portfolios of top 10% (in%)

year	other nonfin. assets	real estate	bus. wealth	equity	liquid assets, bonds	other fin. assets	non housing debt	housing debt	debt-to-asset ratio
1950	0.6	16.0	49.9	21.1	8.0	4.3	30.3	69.7	1.4
1953	0.8	18.4	49.8	20.1	7.2	3.7	32.2	67.8	1.9
1956	0.8	15.9	45.4	27.2	7.1	3.5	17.2	82.8	1.5
1959	0.9	13.3	45.5	30.8	7.4	2.1	16.3	83.7	1.8
1962	0.7	19.6	38.9	30.1	8.5	2.2	7.2	92.8	2.5
1965	0.7	22.3	35.5	33.4	5.4	2.7	11.2	88.8	2.5
1968	0.7	22.8	33.4	32.9	7.5	2.7	10.3	89.7	2.2
1971	0.5	26.4	33.8	27.0	9.4	3.0	9.1	90.9	2.2
1977	0.4	27.8	43.7	15.6	9.6	2.9	12.9	87.1	2.2
1983	2.6	33.6	29.3	13.2	11.6	9.8	35.7	64.3	3.9
1989	3.3	33.8	26.9	8.5	11.5	16.1	27.8	72.2	3.9
1992	2.6	34.1	26.8	10.4	10.1	16.0	17.2	82.8	4.9
1995	3.2	26.6	25.9	14.7	9.8	19.8	17.8	82.2	4.5
1998	2.4	25.0	23.9	18.6	6.7	23.4	22.2	77.8	4.9
2001	2.2	25.3	22.1	17.9	6.7	25.9	17.7	82.3	4.1
2004	2.3	31.5	23.3	15.3	7.6	20.1	15.6	84.4	4.8
2007	1.8	30.3	27.9	15.6	5.9	18.5	11.4	88.6	4.7
2010	2.0	29.3	23.7	14.7	8.5	21.8	11.6	88.4	5.4
2013	1.9	26.1	24.1	16.1	7.1	24.6	10.4	89.6	4.5
2016	1.5	24.2	25.2	19.9	6.3	22.9	17.3	82.7	3.7