Who Benefits from Tax Expenditures on Capital? Evidence on Capital Income and Wealth Concentration*

Emmanuel Saez
UC Berkeley

Gabriel Zucman
UC Berkeley

First version: October 2014
Revised September 2015

Abstract

Tax preferences on capital income and wealth accumulation are a very large component (about half) of all tax expenditures in the US tax system. Tax expenditures on capital include preferential tax rates on realized capital gains and dividends, tax free accumulation of pension and life insurance funds, exclusion of interest from local government bonds, exclusion from income of imputed rent of homeowners combined with the deduction of mortgage interest payments. In this paper, we use tax data from 1979 to 2012 created by the Statistics of Income (SOI) division of the Internal Revenue Service (IRS). We provide systematic evidence on the distribution of wealth and capital income as a first step to improve our understanding of the distributional impact of US tax expenditures on capital. This analysis supplements the evidence provided in Saez and Zucman (2014) using public use tax data. We connect our analysis with the SOI Personal Wealth estimates that are built using estate tax data. Using the estate multiplier technique, we extend the top wealth share estimates of Kopczuk and Saez (2004) up to 2012. We show that, conditional on age and gender, mortality declines with wealth. Furthermore, this gradient has increased sharply since 1979. We discuss how this information could be used to refine Personal Wealth estimates.

*Emmanuel Saez: saez@econ.berkeley.edu; Gabriel Zucman: g.zucman@lse.ac.uk. The tax data were accessed through contract TIRNO-12-P-00374 with the Statistics of Income (SOI) Division at the US Internal Revenue Service. We are very grateful to Victoria Bryant, Barry W. Johnson, and Brian Raub for facilitating the use of the tax data and for their patience in helping us navigate the data. We thank Victoria Bryant, Barry W. Johnson, Janet McCubbin, Brian Raub and Michael Weber for helpful comments.

1 Introduction

Tax expenditures are a very significant component of US federal income tax policy. Burman, Toder, and Geissler (2008) at the Tax Policy Center show that total tax expenditures in 2007 are $702 billion which is about 70% of total federal individual income tax revenue of $1,020 billion collected that year (see Table 1, column (1), Tax Policy Center estimates with AMT in Burman et al. 2008). A very large fraction of such tax expenditures are preferences for capital income and wealth accumulation. The US provides many large tax expenditures for capital income and wealth accumulation. They include in particular preferential tax rates on realized capital gains and dividends ($95 billion in 2007), tax free accumulation of pension and life insurance funds ($145 billion in 2007), the deduction of mortgage interest payments ($92 billion in 2007), exclusion of interest from local government bonds ($12 billion in 2007). The total of such tax expenditures on capital is $345 billion in 2007, or around 50% of total tax expenditures. Note that the Tax Policy Center does not count the exclusion of the imputed rent of homeowners as a tax expenditure even though, conceptually, it is indeed an exclusion of a form of capital income from the income tax base. The Office of Tax Analysis (OTA) at the US Treasury also prepares comprehensive estimates of tax expenditures with projections over the next 10 fiscal years (see US Treasury 2014). OTA estimates are more comprehensive than the Tax Policy Center estimates mentioned above. In particular, OTA estimates also include the imputed rent of homeowners as a tax expenditure. OTA also provides distributional estimates of selected tax expenditure (US Treasury 2014b).

Should capital benefit from preferential treatment through such tax expenditures? On the one hand, tax expenditures on capital could stimulate savings and wealth formation, hereby increasing economic growth. On the other hand, capital income and wealth tend to be concentrated among higher income households so that tax expenditures on capital may reduce the progressivity of the income tax system. This distributional concern looms large at a time when US income concentration is particularly high (Piketty and Saez, 2003).

Therefore, to improve our understanding of the economic consequences of tax expenditures on capital, this paper constructs series on capital income and wealth concentration in the United States. We also construct series of capital income and wealth *composition* to provide the key needed distributional data for the analysis of each of the tax expenditures on capital. For example, the distribution of mortgage debt is relevant for the distributional analysis of the mortgage interest deduction. The distribution of pension fund wealth is relevant for the analysis of the exemption of capital income earned by pension funds. The distribution of corporate equities (held outside pension funds) is relevant for the analysis of the preferential rates on dividends and realized capital gains, etc. This paper builds on the analysis of Saez and Zucman (2014) who use public use tax data to construct series on wealth and capital income concentration and composition since 1913 using the capitalization method. In this paper, we use internal tax data from 1979 to 2012 to supplement the analysis of Saez and Zucman (2014) along five dimensions.

First, we use the SOI individual tax files, available for tax years 1979 to 2012, to extend the estimates of Saez and Zucman (2014) for recent years 2009-2012 (as the public use individual income tax file is currently only available up to 2008). We find that wealth and capital income concentration have grown substantially from 2008 and 2012 implying that the regressivity of tax expenditures on capital has also grown. Importantly, mortgage debt, home real estate, and pension wealth are much more equally distributed than corporate equities assets implying that tax preferences for pensions and home ownership are less regressive than tax preferences for corporate dividends and realized capital gains.

Second, we use information on the age and gender of tax filers to produce series of age and gender composition of top wealth holders from 1979 to 2012. The public use files only have information on whether the tax filers are aged 65 and above (and this information is only available up to year 1995).[1] We find that the relative age of top wealth holders has substantially declined since 1979.

Third, we use information on IRA balances that have been constructed by Bryant and Gober (2013) for years 2004-2011 to calibrate the distribution of pension wealth.

Fourth, we draw links between our analysis and the Personal Wealth Estimates based on estate tax data from the Statistics of Income (SOI) Division at IRS. Johnson (1994, 2011) gathers most of these SOI studies into two comprehensive compendia. More recent SOI studies on wealth are published as SOI working papers posted online or in the SOI bulletin publication. We show that, conditional on age and gender, mortality declines with wealth and that this gradient has increased sharply since 1979. As a result, estimating the wealth distribution from capital income in the sample of decedents and using traditional mortality estimates fails to uncover the increase in wealth concentration from the full population sample. We discuss how to use the statistics we have constructed to improve estate based estimates in the personal wealth studies.

Fifth, we use internal estate tax data for individuals dying in years 1997-2012 matched with their prior year tax return data. Using this matched file, we compute rates of returns for various

[1] After 1995, it can be imperfectly imputed based on the standard deduction (which are higher for tax filers aged 65 and above) and receipt of social security or pension income for itemizers.

asset categories: fixed income claims generating taxable interest income, local government bonds generating tax exempt interest income, publicly traded corporate stock generating dividends and realized capital gains. We also use estate tax data for the period 1976-2012 to extend the top wealth shares series of Kopczuk and Saez (2004) which stopped in 2000 up to 2012, the most recent available year.

This paper builds on a long tradition at SOI investigating the link between income and wealth using matched estate tax data and income tax data including Johnson and Wahl (2004), Rosenmerkel and Wahl (2011), Johnson, Moore, Rosenmerkel (2011), Johnson, Raub, and Newcomb (2013), Bourne and Rosenmerkel (2014). As an important caveat to our analysis, these studies show that the link between income and wealth is much more complex than our simple capitalization method assumes (the reader is referred to Saez and Zucman, 2014 for a detailed discussion of these key methodological issues).

All our estimates are gathered in the companion excel file SOITables.xls that is available online. The first sheet of the excel file called "Explanations" provides a detailed explanations of the information includes in each of the subsequent excel worksheets. These worksheets were directly produced from the underlying tax data using STATA software. Our STATA programs are available upon request to interested researchers who have access to the underlying tax data.

2 Data

Our analysis relies on selected variables from the SOI individual files for tax years 1979 to 2012. The data files include the same variables as the PUF used in the analysis of Saez and Zucman (2014). In addition, the SOI files include gender, year of birth, and year of death of the primary and secondary tax filers. Date of death was obtained by merging the SOI files to the DM1 file (the "Death Master" file) which records birth and death dates of all individuals with valid Social Security Numbers (SSN). The death information was available up to March 2014, allowing us to analyze mortality at a five year horizon up to tax year 2008 (for which we can analyze mortality in the five year period 2009-2013). In addition, we used IRA fair market value data (as of end of the year) for tax years 2004-2011. These files are constructed and described in Bryant and Gober (2013) and can be merged to the SOI files for the corresponding years.

The first step of the data analysis is to construct data files that include the relevant capital income concepts and that homogeneous across years. The second step is to construct individual level wealth variables using our external rates of returns year by year constructed in Saez and Zucman (2014) combining flow of funds aggregate wealth data with published aggregates from individual tax data. The third step is to use the individual level data to construct distributional statistics for various definitions of wealth and capital income. The data and programs are organized to be exactly parallel to the analysis carried out with the PUF in Saez and Zucman (2014).

3 Analysis

3.1 Estimates for 1979-2012

We have used the internal SOI individual tax files available for years 1979 to 2012 to replicate the estimates from Saez and Zucman (2014) made with the public use tax files (PUF). The public use tax files are only available up to 2008. Hence, the SOI files allow us to extend estimates up to 2012. For the period 1979-2008, the PUF and SOI estimates are always very close. They are not exactly identical because the PUF are a subsample of the SOI files.[2] For complete details on the construction of the series, the reader is referred to Saez and Zucman (2014). All these series are reported in a series of tables in the companion excel file. Complete explanations and presentation of each excel sheet is gathered in the front excel sheet called Explanations.

As an alternative, it is possible to use the SOI publicly available tabulations of income and its composition by size of income to provide estimates for 2009-2012. However, because these tabulations are presented by size of income and income rank differs substantially from wealth rank, the resulting estimates are not very precise and differ slightly from the direct estimates using the underlying SOI files. Hence, it is much better to do the estimates using the SOI files. Because the public is particularly interested in current numbers, it is therefore very valuable to be able to use the SOI files to extend the analysis for recent years.[3]

Figure 1 displays the top .1% capital income share (including and excluding realized capital gains). It shows a large increase in the concentration of capital income since the 1970s, with particularly fast increases in recent years. We have also tabulated the distribution of capital income component by component (such as dividend income, interest income, business profits, etc.). These results are presented in detail in Saez and Zucman (2014).

[2]Beginning in 1996, SOI enhanced the disclosure protection procedures applied to high-income records. These procedures resulted in a non-significant downward bias at the top of the income distribution. As a result, we augment the PUF with a synthetic observation constructed so that the PUF matches the published tabulations of the SOI files for the top $10m and above bracket. We have carefully checked using the worksheet sumstats of the excel file SOITables.xls that the aggregate totals are very close for each variable in the SOI files and the corresponding PUF files.

[3]Naturally, having PUF files created and released on a timely basis would also be very valuable as those files are public use and hence can be used by all interested researchers.

Next, we capitalize individual capital income to obtain wealth and estimate top wealth shares. Figure 2 shows the top 10% wealth share (in Panel A) and the top .1% wealth share (in Panel B). Note that the top wealth shares have increased sharply from 2007 to 2012, highlighting the value of being able to use the SOI files to produce high quality estimates for years after 2008.

3.2 Age and Gender Distributions

The PUF have only very limited demographic information. There is no gender information and the only age information is a variable for whether the taxpayers (and his/spouse) is aged 65 or above. This age information is only available up to year 1995. After 1995, it can be imperfectly imputed based on the standard deduction (which are higher for tax filers aged 65 and above) and receipt of social security or pension income for itemizers. We use the SOI files since 1979 to improve the Saez and Zucman (2014) estimates.

We first extend the series on the fraction of wealth coming from taxpayers aged 65 and above up to year 2012 (instead of only up to 1995 from the PUF).

Figure 3 depicts the fraction of wealth held by elderly families for 3 groups : (1) the full population, (2) the bottom 90%, and (3) the top .1%. An elderly family is defined as a tax unit where either the primary filer or the secondary filers (for married tax units) is aged 65 or more. The series covers 1962 to 2012, years for which this information is available. We use the PUF data up to year 1986 and then use the SOI data starting in 1987 (PUF and SOI estimates are very similar in 1986). Three points are worth noting.

First, elderly families held about 1/3 of total wealth in the overall population. This share is very stable from 1962 to 2007 with a slight increase since 2007 to about 36-37%. The share of elderly families in total US families is about 25% in 2010 (up from 18% in 1960).[4] This suggests that elderly families were about twice as wealthy as the average in the 1960s but are now only 40% wealthier than average.

Second, the share of bottom 90% wealth held by elderly households has steadily increased from 12-15% in the 1960s to about 25% in recent years. This trend follows roughly the secular rise in the share of elderly families in the overall population (from 18% in 1960 to 25% in 2010).

Third and most important, the share of top .1% wealth held by elderly households has

[4]US Statistical Abstract 2012, Population Table 62, online at `https://www.census.gov/compendia/statab/2012/tables/12s0062.pdf` for 2010 numbers and `http://www.census.gov/hhes/families/data/households.html` for 1960 numbers. In the Census, elderly families are defined as families with head of household aged 65 or more. This is not exactly the same definition as in the tax data but is very close as, in the vast majority of cases, the head of household is the oldest member of the couple.

followed an opposite trend. In 1962, almost 50% of top .1% wealth was held by elderly household. In 2011, this share has fallen to 36%. Hence, in 1962, top wealth was significantly older than average while today, it is about as old as average. This suggests that the increased concentration of wealth is due primarily to the accumulation of new fortunes out of surging top labor incomes rather than a revival of old accumulated fortunes. This is consistent with the important evidence presented by Edlund and Kopczuk (2009) that the fraction of widows was high in top wealth groups in the 1960s and much lower in the 1990s. This suggests that the fraction of top wealth that is inherited (as opposed to self-made) has decreased since the 1960s.

Next, we provide more details on the age and gender composition of each wealth groups. We compute the fraction of wealth in each group owned by individuals by age groups. For married filers, we assume that wealth is equally split between the two spouses for age group attribution. For consistency, we continue to define wealth groups based on the family tax unit. For each age group, we also compute the fraction of wealth owned by each gender and marital status (single vs. married). These results are reported in the excel file in sheet wealth-baseline-soiage.

3.3 Using IRA Wealth to Calibrate Pension Wealth

We use information on IRA balances that have been constructed by Bryant and Gober (2013) for years 2004-2011 to calibrate the distribution of pension wealth. These statistics are reported in worksheet fracira. Overall IRA wealth in 2011 is around $5.3 Trillion and hence represents about 30% of total pension wealth. We find that about 65% of IRA wealth is held by families with positive pension income on their tax return. About 48% of IRA wealth is held by families with wage earnings above the 50th percentile of the wage earnings distribution. In contrast only about 9% of total IRA wealth is held by families with no pension income or with wage earnings below the 50th percentile of the wage earnings distribution. This justifies our assumption of estimating pension wealth based on pension income and wages above the 50th percentile of the wage earnings distribution and allocating 60% of pension wealth on pension income and 40% of pension wealth on wages above the 50th percentile.

3.4 Link with Personal Wealth Estimates

A large body of work has used the estate multiplier method where estate tax returns are re-weighted by the inverse probability of death (conditional on age and gender) to obtain wealth distributions. SOI presents systematic estimates in its personal wealth studies. Johnson (1994, 2011) provides comprehensive compendia of most recent SOI studies covering years 1982, 1986, 1989, 1992, 1995, 1998, 2001, 2004.[5] Estimates for the full period 1916-2000 have been produced by Kopczuk and Saez (2004) using the comprehensive internal SOI files of estate tax returns since 1916 that are presented in McCubbin (1990).

A number of SOI studies have also compared estate tax data with the SCF data to check the validity of each dataset and potentially estimate the extent of tax avoidance. Scheuren and McCubbin (1994) and Johnson and Woodburn (1994) present such a comparison for years 1983 and 1989 respectively. More recently, Johnson and Moore (2008) provide a more comprehensive comparison covering all years 1989, 1991, 1994, 1997, 2000, 2003 of income reported in the SCF and the income tax returns.

Two SOI studies have estimated estate tax evasion, McCubbin (1994), and Eller et al. (2001). They have used results from Internal Revenue Service audits of estate tax returns for years 1982, and 1992 (respectively) and found relative small tax evasion of about 2-4% for audited returns.[6]

Figure 2, Panel A, shows that our top 10% wealth share estimates are pretty close in both level and trend to the Survey of Consumer Finance (SCF) estimates of Kennickell (2009, 2011). In contrast, as shown in Figure 2, Panel B, there are two large discrepancies between our capitalized income estimates and estate-multiplier estimates of Kopczuk and Saez (2004) or the SCF estimates. First, estate-multiplier estimates deliver a top .1% wealth share that is much lower than our top .1% wealth share in recent decades. Second, estate-multiplier estimates display only a modest increase in the top .1% wealth share since the 1970s. The estate-multiplier method re-weights estate tax returns based on the inverse probability of death. The probability of death is based on mortality tables by age and gender along with a correction to take into account that the wealthy live longer than the average population. In Kopczuk and Saez (2004), the corrective term is obtained from external data on mortality rates of college graduates (a rough proxy for the wealthy) relative to mortality rates of the full population. Kopczuk and Saez (2004) use the same correction factors for all years hereby assuming that the mortality gradient by wealth has not changed overtime. SOI has long recognized and discussed the importance of having accurate mortality rates for the wealthy to estimate accurate wealth distributions from estate tax data (see e.g., the appendix of Raub, 2008). Prior to their 2001 estimates, SOI studies used the National Longitudinal Mortality Survey to compute the mortality differential of the wealthy. Starting with their 2001 estimates, SOI studies use the relative mortality rate

[5] See Schwartz (1994) for year 1982, Schwartz and Johnson (1994) for year 1986, Johnson and Schwartz (1994) for year 1989, Johnson (1997) for years 1992 and 1995, Johnson and Schreiber (2002) for year 1998, Johnson and Raub (2005) for year 2001, and Raub (2008) for year 2004.

[6] Those studies provide a lower bound on estate tax evasion to the extent as audits may fail to uncover all the evaded wealth.

for holders of large dollar value annuity policies obtained from the Society of Actuaries.

Using social security data, Waldron (2007) analyzes both the level and the rate of change in mortality improvement over time by socioeconomic status for male Social Security covered workers. Consistent with our findings presented below, this study finds that the top half of the average relative earnings distribution has experienced faster mortality improvement than has the bottom half. More precisely, male Social Security, covered workers born in 1941 who had average relative earnings in the top half of the earnings distribution and who lived to age 60 would be expected to live 5.8 more years than their counterparts in the bottom half. For the 1912 cohort, the corresponding difference was only 1.2 years (instead of 5.8 years).[7]

Using the SOI individual income tax samples for tax years 1979 to 2008 which have information on age and date of death from the DM1 file, we can estimate the mortality rates by age, gender, and wealth class as reported in Table 1. The table reports mortality statistics by age, gender, wealth group, and time period. Cols. [1]-[5] is for men, cols. [6]-[10] is for women. These estimates are computed using SOI individual annual files since 1979 combined with the death master DM1 file. Mortality rates are computed on a 5-year horizon (dividing by 5 to obtain annualized mortality rates) and by 5 year age groups. To reduce noise, this table reports mortality statistics aggregated by larger age groups (30-49, 50-64, 65-79, 80+) using the raw data and always using the population weights for the period 2004-2008 of the corresponding gender and wealth group (to avoid bias due to changes in the age distribution). Cols. [1] and [6] report the average annual mortality rate among all tax filers (which represent 94% of the full population on average). Cols. [2] and [7] report the mortality differential for men and women used by Kopczuk and Saez (2004) in their estate multiplier method (the aggregation by age groups uses the top 10% wealthiest age and gender distribution in the period 2004-2008). This mortality differential is estimated in Brown, Liebman, Pollet (2002) and measures the mortality rate of white college graduates relative to the full population in years 1979-1985 using the National Longitudinal Mortality Survey. Kopczuk and Saez (2004) use this mortality differential for all years in their estate multiplier estimates. Cols. [3]-[5] and [8]-[10] report the relative mortality for top 10%, top 5%, and top 1% wealth holders for men and women. Wealth is measured at the family tax unit level.

We illustrate the key features of these statistics on Figures 4. The top two panels report the

[7]A number of earlier studies in the United States have found evidence of widening of mortality differential by socio-economy status (either life-time earnings, or educational achievement). See e.g., Feldman et al., Duleep (1989), Pappas et al. (1993), Waldron (2004). To our knowledge, we are the first to estimate trends in mortality differentials by wealth in the recent period.

mortality rate of males (left panel) and females (right panel) by wealth group and age groups relative to the full population in the same gender and age group. The Kopczuk-Saez series is the mortality rate of college goers used to correct the estate multipliers in Kopczuk and Saez (2004). Panel A shows that, for males, the mortality correction for the top 10% overall is very close to the mortality correction used by Kopczuk and Saez (2004). However, this is a clear mortality gradient within the top 10%. The top 10% live less long than the top 1% who in turn live less long than the top .1%. Computations are based on SOI tax data for years 1999 to 2008. The bottom panels plot the mortality rates by wealth group relative to the full population for men aged 65 to 79 over time from 1979 to 2008 (averaged by groups of 5 years to reduce noise) for men (left panel) and women (right panel) separately. Both graphs show an increasing gradient of mortality with wealth over time. The trend is especially pronounced for men. This implies that the bias of the estate multiplier method used in Kopczuk and Saez (2004) is going to grow over time.

To test for such a bias directly, we estimate top wealth shares using our capitalization method for the subset of individuals who die the following year. The mimics the sample of decedents used by the estate multiplier method. We re-weight individuals using the same weights as the estate multiplier method of Kopczuk and Saez (2004): the weight is equal to the inverse of the probability of death by age×gender×year (obtained from population wide mortality tables) times the socio-economic mortality correction factor (based on age and gender) from Kopczuk and Saez (2004). Note that this mortality correction factor is invariant by year. Therefore, such series isolate the level and trend effects of the mortality gradient by wealth class. Figure 5 depicts the top 1% wealth shares based on the full sample and capitalization method, the top 1% wealth shares based on the subsample of decedents and capitalization method, and the top 1% wealth shares obtained by Kopczuk and Saez (2004) for years 1916-2000, Johnson and Raub (2005) for year 2001, and Raub (2008) for year 2004 from estate tax data. For the subsample of decedents, we re-weight individuals using the same weights as the estate multiplier method of Kopczuk and Saez (2004): the weight is equal to the inverse of the probability of death by age×gender×year (obtained from population wide mortality tables) times the socio-economic mortality correction factor (based on age and gender) from Kopczuk and Saez (2004). Note that this mortality correction factor is invariant by year. Married filers carry a weight of one half when one spouse dies. Therefore, such series isolate the level and trend effects of the mortality gradient by wealth class. The fact that the series for decedents is flat and very similar to the estate based series of Kopczuk and Saez (2004) suggests that the mortality gradients and trends documented in the previous figure can fully explain the discrepancy between the Kopczuk and Saez (2004) estimates and our estimates using the capitalization method.

Overall the results from Figures 4 and 5 suggest that the estate tax multiplier method fails to capture the increase in top wealth shares because of a sharply widening mortality gradient by wealth class.

Therefore, our capitalization method should be useful to obtain more precise mortality correction factors by wealth class to improve estate multiplier estimates. Hence, we see the two approaches as complementary. Note also that, since 2010, the estate tax exemption threshold has been above $5 million so that estate tax data will cover only about .25% of the wealth distribution. Hence, the capitalization method appears as the only way to have long run, yearly series that cover the full distribution including the very top groups.

Linked estate and income tax data. Linked estate and income tax data can be used to compute rates of returns by asset class and wealth group to test the validity of the capitalization method, which assumes constant rates of returns by asset class and wealth group. We proceed in two steps.

First, we use publicly available SOI tabulations of matched estate-income returns for estates filed in 2008, typically 2007 decedents matched to their 2006 income. As shown in the top Panel of Figure 6, within-asset-class returns appear constant across wealth groups. In each estate tax bracket, the interest yield is about 3% and the dividend yield close to 3.5%. When including realized capital gains, the equity return is about 8-9% across the board.[8] Although taxable rates of returns vary across individuals, they are similar across wealth groups.

Second, we use the internal SOI sample files to conduct a systematic, micro analysis of rates of return at death for each year from 1996 to 2011. We match the estate tax returns of non-married decedents to their prior-year income tax returns. We again find that rates of return are similar across the wealth distribution. As shown in the bottom Panel of Figure 6, the interest rate on taxable bonds and deposits does not vary much with wealth over the 1996-2011 period. In 1997, for example, the interest rate is 3.9% on aggregate, and between 4.1% and 4.3% for all groups of estate tax payers ranging from $0.5–1 million to more than $20 million. As reported in the accompanying tables, we find similarly negligible returns differentials for tax-exempt municipal bonds. The one exception is that we find a modest taxable interest rate

[8]This evidence is consistent with the more detailed analysis by Johnson, Raub, and Newcomb (2013) using micro estate tax data of 2007 decedents matched to 2006 income tax returns. If anything, Johnson, Raub, and Newcomb (2013) find slightly decreasing rates of returns for some asset classes (see their Figure 2), suggesting that our capitalization method might actually slightly understate wealth concentration in 2006.

premium for estates above $20 million in 2003, 2005 and since 2008.

Understanding the discrepancy with estate based estimates. As shown in the top panel of Figure 7, from 1916 to 1976 the estate-based top 0.1% wealth share is remarkably similar to ours in both level and trend. The similarity despite different sources and methods gives credibility to the finding that wealth concentration declined a lot during the first half of the 20th century. However, there is a large discrepancy between the two sources after 1976: we find a sharp increase in wealth concentration, while estate data display only a modest rise. In particular, extending the Kopczuk and Saez (2004) estimates which stop in 2000 up to year 2012 (using exactly the same methodology) shows that estate based estimates remain almost flat. Note also that estate tax estimates can potentially be fairly sensitive to the death of a single very rich person. For example, Steve Jobs died at age 56 in 2011 with a fortune estimated by Forbes magazine to be $7 billion in 2011. The weight for a male aged 56 in the estate multiplier technique is 198.6 so that a wealth of $7 billion would represent $1.4 trillion, or 3.1% of the total wealth denominator for this year. This is enough to explain the 3 point spike in 2011 visible in the extended Kopczuk and Saez (2014) series.[9] In principle, this gap could partly owe to differences in the unit of observation, which is the individual in estates tax data and the tax unit in capitalized income series. In the associated tables, we report individual-level top wealth shares obtained by capitalizing the income of tax units and splitting the wealth of married couples equally; moving to the individual unit makes a negligible difference. How then can we explain the gap between estate-mutlplier estimates and ours?

To cast light on this issue, we run a simple test: we apply the estates multiplier technique not to wealth but to income. Using internal SOI tax return sample files since 1979, we match the estates tax return of decedents to their prior-year income tax return. We then compute the distribution of capital income at death weighting each observation by the Kopczuk and Saez (2004) inverse mortality rates. If the estate multiplier technique worked well, the distribution of capital income in the weighted decedent sample should be similar as that in the living population. However, as the bottom panel of Figure 7 shows, it is not. In 1976, the top 0.1% capital income share is about 15% in both weighted estate-income data and in the overall population. But according to the estate multiplier method, the concentration of capital income has barely increased since 1976, while in actual facts it has surged. One could argue that individuals should

[9]Naturally, this computation is purely illustrative. It is possible that the wealth estimated by Forbes magazine for Steve Jobs could be off by a significant margin relative to his estate, for example if his wealth is owned jointly with his spouse. Raub, Johnson, and Newcomb (2011) carry out a systematic comparison of Forbes 400 decedents reported estates and their wealth estimated by Forbes magazine. They show that, on average, the values reported for tax purposes are approximately half those estimated by Forbes.

not realize capital gains at the end of their life in order to take advantage of the tax free step-up of basis at death. However, this would bias the level and not the trend as this tax incentive has always existed. Furthermore, the discrepancy is similar when excluding realized capital gains from income. One could argue again that some components of capital income reflect labor income (as in the case of business profits for active business owners). However, the discrepancy remains the same when focusing on "passive" capital income only, namely dividends, interest, and rents. These components are least likely to be contaminated by labor income. Therefore, we conclude that the weighted decedent sample has become less and less representative of the living population, explaining why the estate multiplier technique fails to capture the surge in top wealth shares.

4 Conclusion

The series of wealth and capital income created in this study are the key input needed to carry out a distributional analysis of tax expenditures on capital. The analysis carried out in this paper could be further extended using internal data along several promising directions.

First, wealth estimates could be refined as follows. The value of real estate could be capitalized with state specific capitalization factors (instead of a single national capitalization factors) proportional to the effective property tax rates on residential real estate in each state.[10] More ambitiously, the value of homes could be estimated using address information in tax data matched to existing third-party home price databases. Mortgage debt of non-itemizers could be estimated using 1098 information returns. The value of pension wealth could be estimated more precisely using several pieces of information. The value of all IRAs are available at the individual micro-level and IRAs represent about 30% of all pension wealth. The value of employer Defined Benefits and Defined Contribution pensions could be estimated from employer level information as well as past employment and contributions of the individual obtained from W2 forms. The value of businesses (such as partnerships and S-corporations) could be estimated by matching individual returns of owners with the corresponding business tax return balance sheets.

Second, a key distinction is whether wealth comes from inheritance vs. savings from work. The data we have constructed could be used to estimate flows of inheritance by wealth group. This information could then be used to estimate the inheritance flow which can be used to estimate the share of wealth coming from inheritance as shown in Piketty and Zucman (2014).

[10]Such state specific property tax rates could be estimated using US census bureau data as done for example by the Tax Foundation at at `http://taxfoundation.org/article/property-taxes-owner-occupied-housing-state-2004-2009`.

References

Bourne, Jenny and Lisa Rosenmerkel. 2014. "Over the Top: How Tax Returns Show that the Very Rich Are Different from You and Me," Statistics of Income, IRS Working Paper.

Bryant, Victoria L. and Jon Gober. 2013. "Accumulation and Distribution of Individual Retirement Arrangements, 2010", *Internal Revenue Service: Statistics of Income Bulletin*, Fall 2013, 1–18.

Burman, Leonard, Eric Toder, and Christopher Geissler. 2008. "How Big Are Total Individual Income Tax Expenditures, and Who Benefits from Them?", *American Economic Review* 98(2), 79–83.

Duleep, Harriet Orcutt. 1989. "Measuring socioeconomic mortality differentials over time." *Demography* 26(2): 345–351.

Edlund, Lena and Wojciech Kopczuk. 2009. "Women, Wealth and Mobility," *American Economic Review*, 99(1), 146-78.

Feldman, Jacob J., Deane M. Makuc, Joel C. Kleinman, and Joan Cornoni-Huntley. 1989. "National trends in education differentials in mortality." *American Journal of Epidemiology* 129(5): 919–933.

Johnson, Barry W., ed. 1994. *Compendium of Federal Estate Tax and Personal Wealth Studies* Department of Treasury, Internal Revenue Service, Publication 1773 (4-94).

Johnson, Barry W. 1997. "Personal Wealth, 1992-1995", *Internal Revenue Service: Statistics of Income Bulletin*, Winter 1997-1998, 70–95.

Johnson, Barry W., ed. 2011. *Compendium of Federal Estate Tax and Personal Wealth Studies*, Volume 2, Department of Treasury, Internal Revenue Service.

Johnson, Barry W. and Kevin Moore 2008. "Differences in Income Estimates Derived from Survey and Tax Data," Statistics of Income, IRS Working Paper.

Johnson, Barry W., Kevin Moore, and Lisa M. Rosenmerkel 2011. "The Income-Wealth Paradox: Connections Between Realized Income and Wealth Among Americas Aging Top Wealth-Holders", in Johnson, B. W. ed., *Compendium of Federal Estate Tax and Personal Wealth Studies*, Volume 2, Department of Treasury, Internal Revenue Service.

Johnson, Barry W. and Brian Raub 2005. "Personal Wealth, 2001", *Internal Revenue Service: Statistics of Income Bulletin*, Winter 2005-2006, 120–146.

Johnson, Barry W., Brian Raub, and Joseph Newcomb 2013. "A New Look at the Income-Wealth Connection for Americas Wealthiest Decedents", Statistics of Income, IRS Working Paper.

Johnson, Barry W. and Lisa M. Schreiber 2002. "Personal Wealth, 1998", *Internal Revenue Service: Statistics of Income Bulletin*, Winter 2002-2003, 87–115.

Johnson, Barry W. and Marvin Schwartz. 1994. "Estimates of Personal Wealth, 1989," in Johnson, B. W. ed., *Compendium of Federal Estate Tax and Personal Wealth Studies* Department of Treasury, Internal Revenue Service, Publication 1773 (4-94), 287–304.

Johnson, Barry and Jenny Wahl. 2004. "The Mismeasurement of Man's Well-Being: Refining Realized Income Measures with Wealth, Portfolio, and Mortality Information," Statistics of Income, IRS Working Paper.

Johnson, Barry W. and R. Louise Woodburn. 1994. "The Estate Multiplier Technique: Recent Improvements for 1989," in Johnson, B. W. ed., *Compendium of Federal Estate Tax and Personal Wealth Studies* Department of Treasury, Internal Revenue Service, Publication 1773 (4-94), 391–400.

Kennickell, Arthur B. 2009. "Ponds and Streams: Wealth and Income in the US, 1989 to 2007," FEDS Working Paper, Board of Governors of the Federal Reserve System.

Kennickell, Arthur. 2011. "Tossed and Turned: Wealth Dynamics of US Households 2007-2009", FEDS Working Paper, Board of Governors of the Federal Reserve System.

Kopczuk, Wojciech. 2007. "Bequest and Tax Planning: Evidence from Estate Tax Returns," *Quarterly Journal of Economics*, 122(4), 1801-1854.

Kopczuk, Wojciech and Emmanuel Saez. 2004. "Top Wealth Shares in the United States, 1916-2000: Evidence from Estate Tax Returns," *National Tax Journal*, 57(2), Part 2, 445-487.

McCubbin, Janet G. 1990. "The Intergenerational Wealth Study: Basic Estate Data, 1916-1945," *Internal Revenue Service: Statistics of Income Bulletin*, Spring 1990.

McCubbin, Janet G. 1994. "Improving Wealth Estimates Derived from Estate Tax Data", in Johnson, B. W. ed., *Compendium of Federal Estate Tax and Personal Wealth Studies* Department of Treasury, Internal Revenue Service, Publication 1773 (4-94).

Pappas, Gregory, Susan Queen, Wilbur Hadden, and Gail Fisher. 1993. "The increasing disparity in mortality between socioeconomic groups in the United States, 1960 and 1986." *New England Journal of Medicine* 329(2): 103–109.

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

Piketty, Thomas, and Gabriel Zucman. 2014."Wealth and Inheritance in the Long Run," in Anthony B. Atkinson and Francois Bourguignon, eds., *Handbook of Income Distribution,* Volume 2, Amsterdam; New York: Elsevier/North Holland.

Raub, Brian G. 2008. "Personal Wealth, 2004", *Internal Revenue Service: Statistics of Income Bulletin*, Fall 2008, 282–308.

Raub, Brian G., Barry W. Johnson, and Joseph Newcomb 2011. "A Comparison of Wealth Estimates for Americas Wealthiest Decedents Using Tax Data and Data from the Forbes 400", in Johnson, Barry W., ed. *Compendium of Federal Estate Tax and Personal Wealth Studies*, Volume 2, Department of Treasury, Internal Revenue Service.

Rosenmerkel, Lisa and Jenny Wahl. 2011. "Crossing the Bar: Predicting Wealth from Income and Estate Tax Records," in Johnson, B. W. ed., *Compendium of Federal Estate Tax and Personal Wealth Studies*, Volume 2, Department of Treasury, Internal Revenue Service.

Saez, Emmanuel and Gabriel Zucman. 2015 "Wealth Inequality in the United States since 1913: Evidence from Capitalized Income Tax Data," NBER Working Paper No. 20625, October 2014, revised September 2015.

Scheuren, Fritz and Janet McCubbin. 1994. "Piecing Together Personal Wealth Distributions." in Johnson, B. W. ed., *Compendium of Federal Estate Tax and Personal Wealth Studies* Department of Treasury, Internal Revenue Service, Publication 1773 (4-94), 371–390.

Schwartz, Marvin and Barry W. Johnson. 1994. "Estimates of Personal Wealth, 1986," in Johnson, B. W. ed., *Compendium of Federal Estate Tax and Personal Wealth Studies* Department of Treasury, Internal Revenue Service, Publication 1773 (4-94). 255–270.

U.S. Treasury Department, Internal Revenue Service. 1916-2012. *Statistics of Income: Individual Income Tax Return,* annual from 1916 to 2012, Washington, D.C.

U.S. Treasury Department, Internal Revenue Service. "Statistics of Income: Estate and Gift Tax Returns," various years. Washington, D.C.

U.S. Treasury Department, Internal Revenue Service. "Statistics of Income: Individual and Fiduciary Income Tax Returns," various years. Washington, D.C.

U.S. Treasury Department, Office of Tax Analysis. 2014. "Tax Expenditures." Washington, D.C.

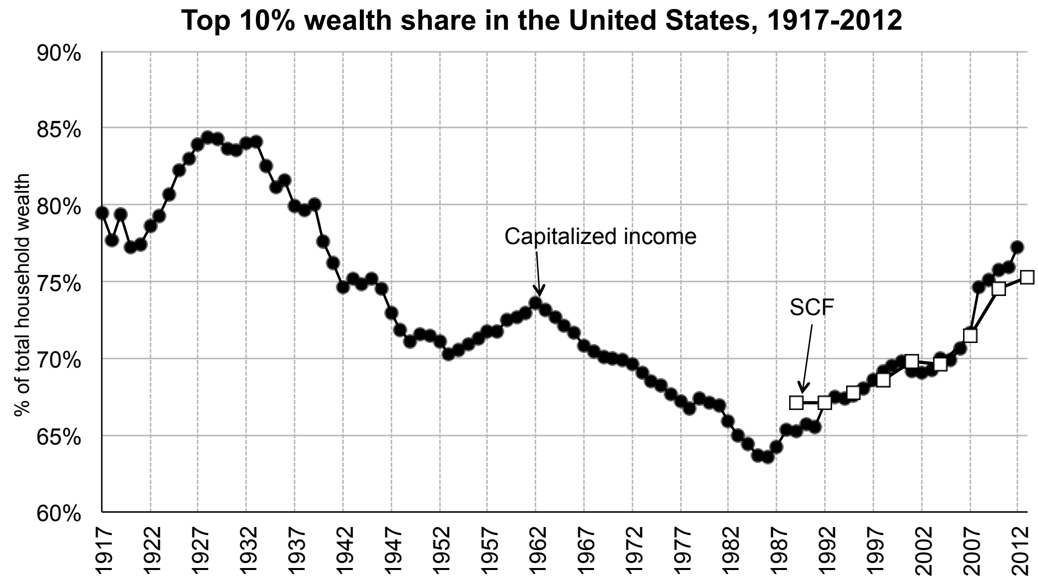
U.S. Treasury Department, Office of Tax Analysis. 2014b. "Distribution of Selected Tax Expenditures." Washington, D.C

Waldron, Hilary. 2004. "Heterogeneity in health and mortality risk among early retiree men." ORES Working Paper Series, No. 105. Social Security Administration, Office of Policy, Office of Research, Evaluation, and Statistics, Washington, DC (May).

Waldron, Hilary. 2007. "Trends in Mortality Differentials and Life Expectancy for Male Social Security Covered Workers, by Average Relative Earnings", ORES Working Paper No. 108, Social Security Administration, Office of Policy, Office of Research, Evaluation, and Statistics, Washington, DC (May).

Figure 1: The Top 0.1% Taxable Capital Income Share in the United States, 1962-2012

Notes: The figure plots the top 0.1% taxable capital income share in the United States from 1962 to 2012. Taxable capital income includes dividends, taxable interest, positive rents, estate and trust income, as well as the positive profits of S-corporations, sole proprietorships and partnerships (negative profits and rental income are disregarded). It excludes tax exempt interest paid by state and local bonds (munis). The top curve includes taxable capital gains. The unit is the family (either a single person aged 20 or above or a married couple, in both cases with children dependents if any). Complete details on definitions and computations are in the excel appendix file.



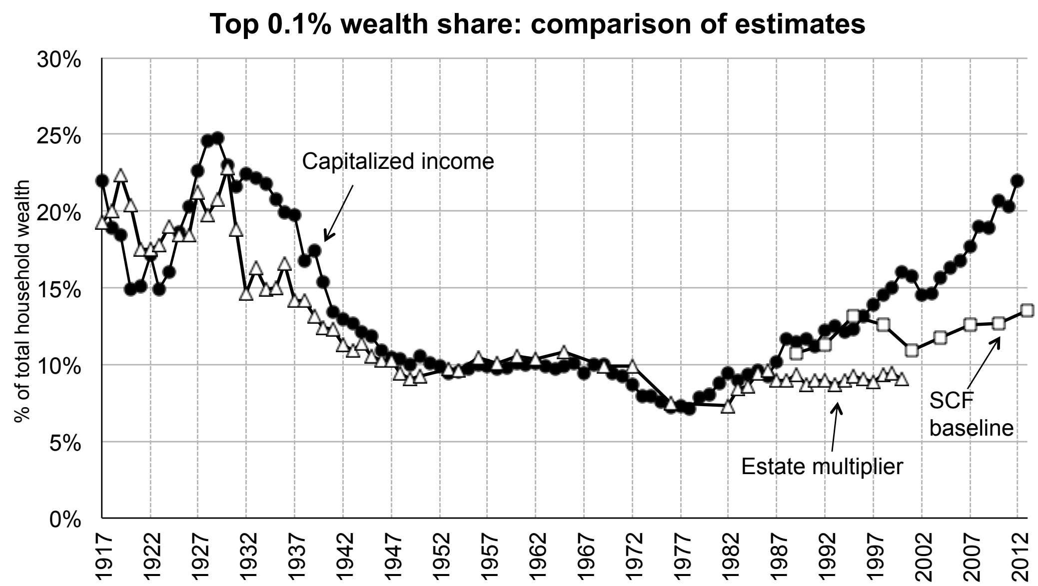


Figure 2: Top Wealth Shares Estimated with Capitalization Method and Earlier Estimates

Notes: This Figure compares our top wealth share estimates with earlier top wealth shares estimates from the Survey of Consumer Finances (SCF) by Kennickell (2009, 2011) and estimates using estate tax returns by Kopczuk and Saez (2004). Panel A focuses on the top 10% and panel B on the top .1%. Complete details on definitions and computations are in the excel appendix file.

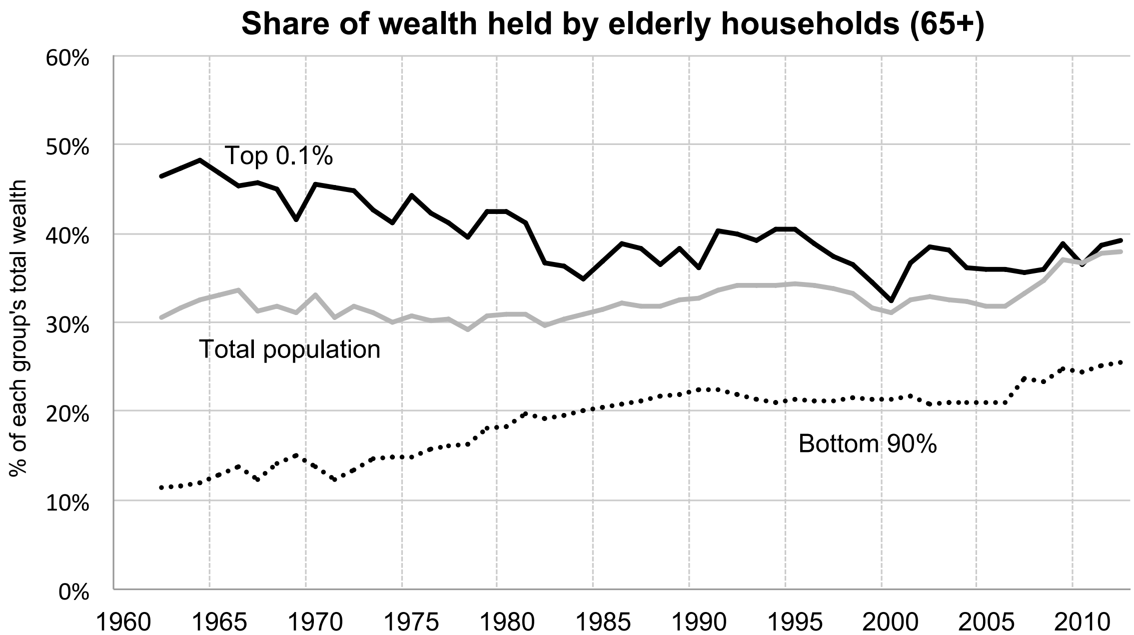


Figure 3: Share of Wealth Held by Elderly Households, 1962-2012

Notes: The figure depicts the fraction of wealth held by elderly families for 3 groups : (1) the full population, (2) the bottom 90%, and (3) the top .1%. An elderly family is defined as a tax unit where either the primary filer or the secondary filer (for married tax units) is aged 65 or more. The series covers 1962 to 2012. Series before 1987 use the Public Use Files. Series after 1986 use the SOI files. Complete details on definitions and computations are in the excel appendix file.

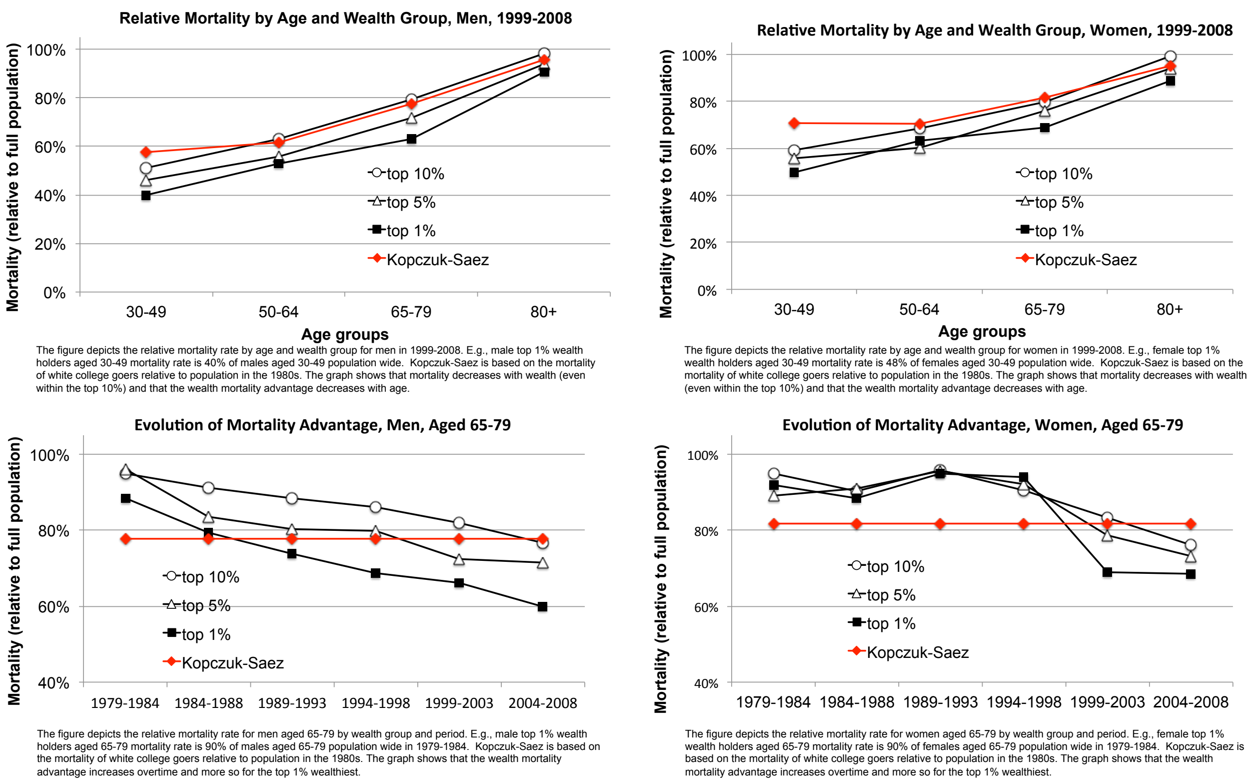


Figure 4: Mortality Gradients by Wealth Group

Notes: The top two panels report the mortality rate of males (left panel) and females (right panel) by wealth group and age groups relative to the full population in the same gender and age group. The Kopczuk-Saez series is the mortality rate of college goers used to correct the estate multipliers in Kopczuk and Saez (2004). Panel A shows that, for males, the mortality correction for the top 10% overall is very close to the mortality correction used by Kopczuk and Saez (2004). However, this is a clear mortality gradient within the top 10%. The top 10% live less long than the top 1% who in turn live less long than the top .1%. Computations are based on SOI tax data for years 1999 to 2008. The bottom panels plot the mortality rates by wealth group relative to the full population for men aged 65 to 79 over time from 1979 to 2008 (averaged by groups of 5 years to reduce noise) for men (left panel) and women (right panel) separately. Both graphs show an increasing gradient of mortality with wealth over time. The trend is especially pronounced for men. This implies that the bias of the estate multiplier method used in Kopczuk and Saez (2004) is going to grow over time. Complete details on definitions and computations are in the excel appendix file.

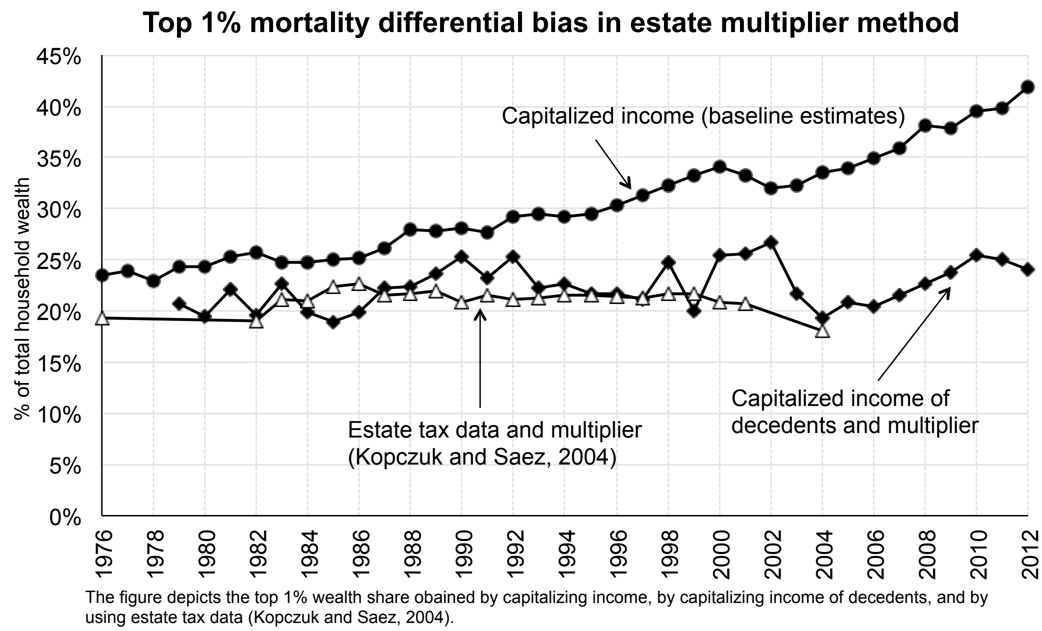
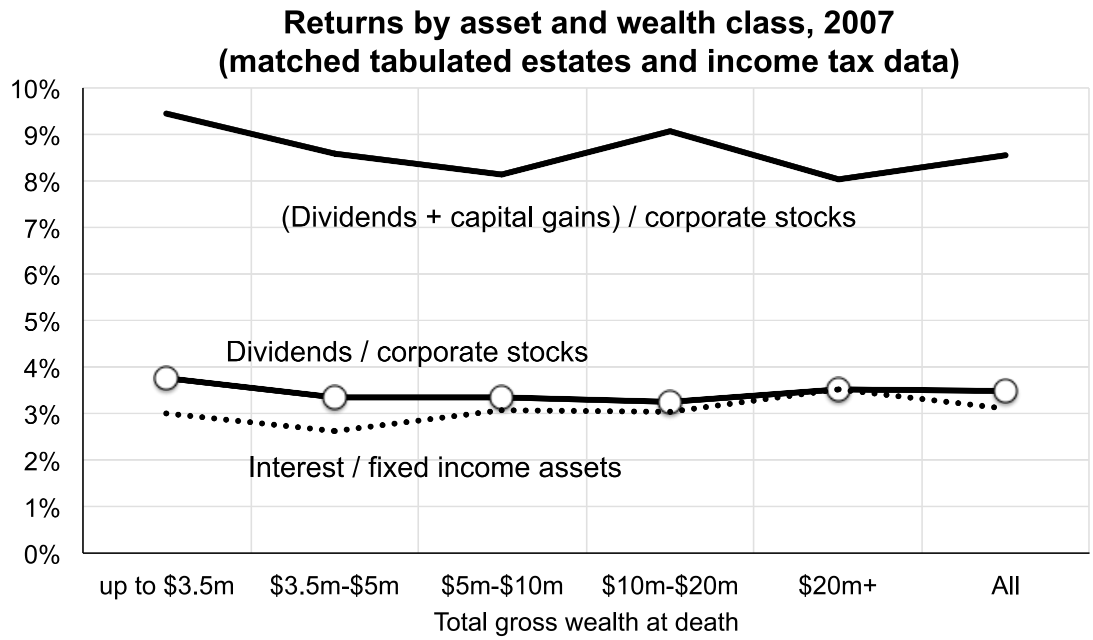
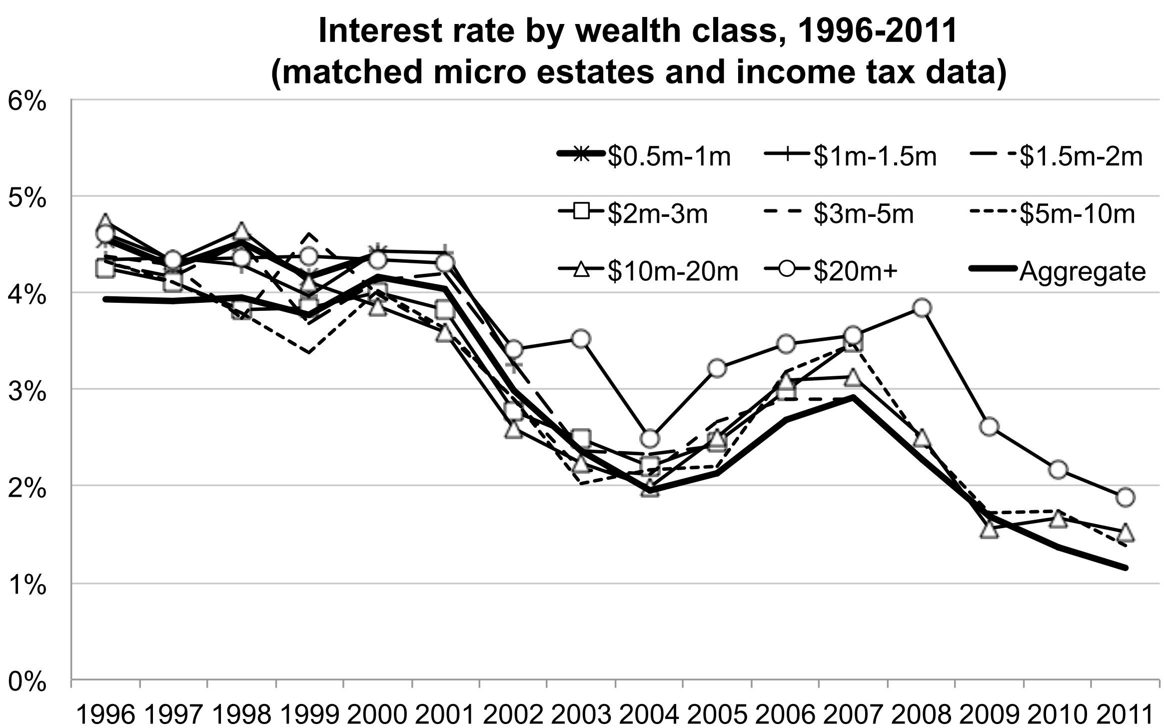


Figure 5: Top 1% Wealth Shares: Decedents vs. Full Sample

Notes: The figure depicts the top 1% wealth shares based on the full sample and capitalization method, the top 1% wealth shares based on the subsample of decedents and capitalization method, and the top 1% wealth shares obtained by Kopczuk and Saez (2004) for years 1916-2000, Johnson and Raub (2005) for year 2001, and Raub (2008) for year 2004 from estate tax data. For the subsample of decedents, we re-weight individuals using the same weights as the estate multiplier method of Kopczuk and Saez (2004): the weight is equal to the inverse of the probability of death by age×gender×year (obtained from population wide mortality tables) times the socio-economic mortality correction factor (based on age and gender) from Kopczuk and Saez (2004). Note that this mortality correction factor is invariant by year. Married filers carry a weight of one half when one spouse dies. Therefore, such series isolate the level and trend effects of the mortality gradient by wealth class. The fact that the series for decedents is flat and very similar to the estate based series of Kopczuk and Saez (2004) suggests that the mortality gradients and trends documented in the previous figure can fully explain the discrepancy between the Kopczuk and Saez (2004) estimates and our estimates using the capitalization method. Complete details on definitions and computations are in the excel appendix file.



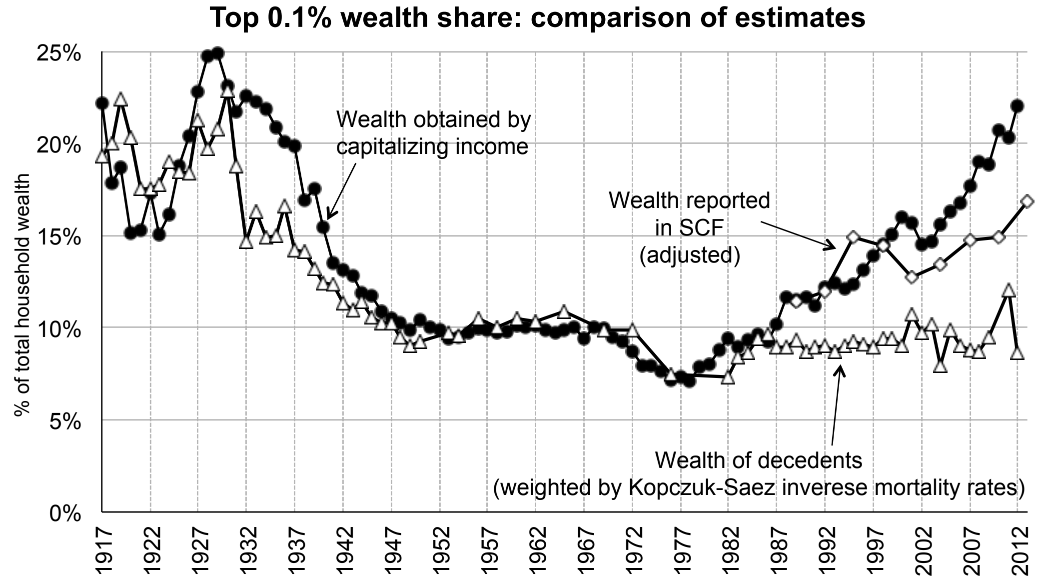
The figure reports returns for various assets classes by size of gross worth using matched estate and prior year income tax data for 2008 estate tax filers (mostly 2007 decedents), excluding joint filers. Source: Appendix Table C6.



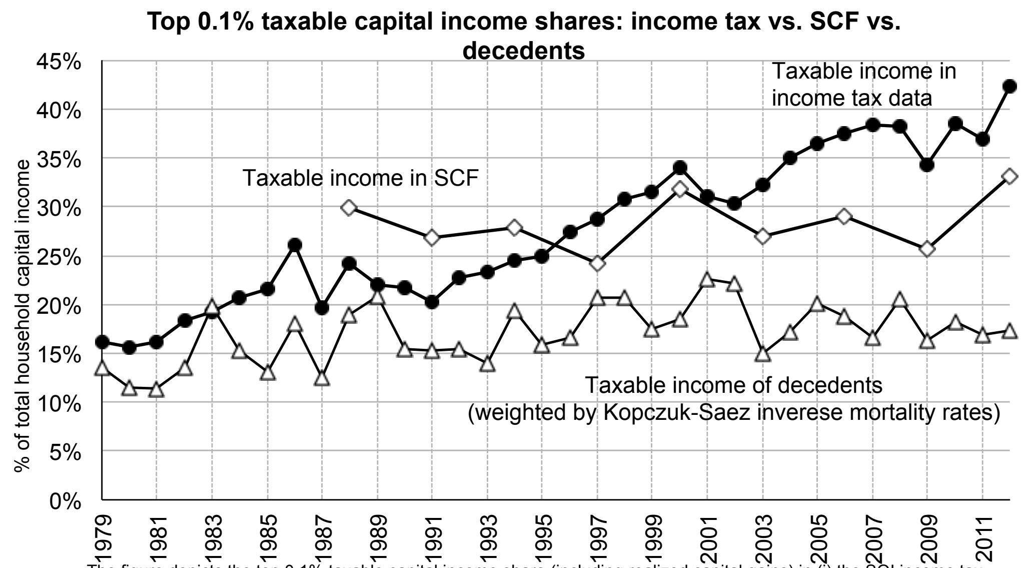
The figure shows how the interest rate varies across the distribution of wealth at death using matched estate and prior year income tax data for 1997 to 2012 decedents, excluding joint filers. Source: Appendix Table C6b.

Figure 6: Taxables Rates of Returns by Wealth

Notes: The figure displays how taxable rates of returns vary across the distribution of wealth at death, using estate tax returns matched to prior year income tax returns of non-married filers. Individuals are ranked by their size of wealth at death (gross wealth in the top panel and net wealth in the bottom panel). The top panel uses published tabulated data for estates filed in 2008 (typically 2007 decedents). The bottom panel uses internal SOI estate tax returns matched to prior-year income tax returns; the year denotes the income-tax year (for instance, 1996 refers to estates for decedents in 1997 matched to the decedent's 1996 income tax return). In all cases, within-asset class returns appear to be fairly stable across wealth groups.



The figure depicts the top 0.1% wealth share obained by capitalizing income, by using the SCF (adjusted to move to tax-unit level, include Forbes 400, ant match Financial Accounts totals), and by using estate tax data. Source: Appendix C4 and C4b.



The figure depicts the top 0.1% taxable capital income share (including realized capital gains) in (i) the SOI income tax data; (ii) the sample of decedents weighted using the Kopczuk-Saez (2004) estate mutiplier weights; (iii) the SCF. For both the SCF and SOI income tax data, fractiles are defined relative to the total number of tax units.

Figure 7: Reconciling Capitalized Top Wealth Shares with Other Estimates

Notes: The top panel compares our top 0.1% wealth share estimates with top wealth shares estimates from using estate tax returns (Kopczuk-Saez (2004) for 1917-2000, and our own computations applying the same methodology as Kopczuk-Saez (2004) for 1981, 2001-2009, 2011-2) and the Survey of Consumer Finances (SCF). To improve comparability, starting from the SCF baseline estimates of Kennickell (2009, 2011), we adjusted SCF data by: (1) defining fractiles relative to total tax units instead of households, (2) adjusting individual wealth components to match household balance sheet totals asset class by asset class, (3) adding back the Forbes 400 that are excluded by design from the SCF. The bottom panel compares the top 0.1% capital income shares estimates from the SOI income tax data, the SCF, and the estate tax (where estates are matched to income tax returns, and decedents are weighted by the inverse of their mortality rate, using the weights of Kopczuk and Saez (2004)). In the three cases, we use the same definition of capital income (as the SCF reports income following the lines of the income tax return). Namely, capital income is the sum of (taxable) interest income, dividends, realized capital gains, profits from sole proprietorships, partnerships and S-corporations, rents, royalties (schedule C and schedule E income). For both the SOI income tax data and the SCF, fractiles are defined relative to the total number of tax units.

Back to index

Table 1: Relative Mortality by Wealth Group, Age, Gender, and Time Period

	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]
	Men					Women				
	Population wide annual mortality rate	Relative mortality				Population wide annual mortality rate	Relative mortality			
		Kopczuk-Saez	Top 10%	Top 5%	Top 1%		Kopczuk-Saez	Top 10%	Top 5%	Top 1%
Period 1979-1983										
Age 30-49	0.27%	58%	73%	72%	63%	0.15%	71%	64%	57%	64%
Age 50-64	1.30%	62%	74%	69%	61%	0.65%	70%	87%	78%	71%
Age 65-79	3.85%	78%	95%	96%	88%	2.14%	82%	95%	89%	92%
Age 80+	7.62%	96%	101%	99%	103%	6.88%	95%	113%	120%	115%
Period 1984-1988										
Age 30-49	0.26%	58%	81%	80%	55%	0.15%	71%	78%	66%	45%
Age 50-64	1.17%	62%	76%	71%	69%	0.60%	70%	89%	79%	80%
Age 65-79	3.70%	78%	91%	83%	79%	2.20%	82%	90%	91%	88%
Age 80+	8.73%	96%	105%	106%	88%	7.00%	95%	91%	92%	87%
Period 1994-1993										
Age 30-49	0.28%	58%	60%	59%	57%	0.13%	71%	71%	66%	73%
Age 50-64	1.04%	62%	75%	71%	53%	0.60%	70%	94%	93%	79%
Age 65-79	3.59%	78%	88%	80%	74%	2.25%	82%	96%	96%	95%
Age 80+	9.04%	96%	101%	100%	82%	6.95%	95%	102%	101%	105%
Period 1994-1998										
Age 30-49	0.23%	58%	62%	55%	76%	0.14%	71%	56%	61%	64%
Age 50-64	0.90%	62%	77%	73%	67%	0.59%	70%	86%	81%	64%
Age 65-79	3.25%	78%	86%	80%	69%	2.06%	82%	90%	92%	94%
Age 80+	8.78%	96%	103%	104%	94%	6.96%	95%	99%	93%	89%
Period 1999-2003										
Age 30-49	0.22%	58%	50%	48%	27%	0.14%	71%	67%	55%	59%
Age 50-64	0.84%	62%	65%	59%	64%	0.54%	70%	71%	63%	55%
Age 65-79	3.04%	78%	82%	72%	66%	1.99%	82%	83%	79%	69%
Age 80+	8.73%	96%	100%	97%	89%	7.36%	95%	102%	96%	86%
Period 2004-2008										
Age 30-49	0.20%	58%	52%	44%	53%	0.13%	71%	51%	57%	40%
Age 50-64	0.81%	62%	61%	53%	43%	0.48%	70%	67%	57%	71%
Age 65-79	2.76%	78%	77%	71%	60%	1.99%	82%	76%	73%	69%
Age 80+	8.47%	96%	97%	91%	92%	7.22%	95%	97%	92%	91%

Notes: This table reports mortality statistics by age, gender, wealth group, and time period. Cols. [1]-[5] is for men, cols. [6]-[10] is for women. These estimates are computed using Statistic of Income individual annual files since 1979 combined with the death master DM1 file). Mortality rates are computed on a 5-year horizon (dividing by 5 to obtain annualized mortality rates) and by 5 year age groups. The raw data is presented in the sheet mortality_raw. To reduce noise, this table reports mortality statistics aggregated by larger age groups (30-49, 50-64, 65-79, 80+) using the raw data and always using the population weights for the period 2004-2008 of the corresponding gender and wealth group (to avoid bias due to changes in the age distribution). Cols. [1] and [6] report the average annual mortality rate among all tax filers (which represent 94% of the full population on average). Cols. [2] and [7] report the mortality differential for men and women used by Kopczuk and Saez (2004) in their estate multiplier method (the aggregation by age groups uses the top 10% wealthiest age and gender distribution in the period 2004-2008). This mortality differential is estimated in Brown, Liebman, Pollet (2002) and measures the mortality rate of white college graduates relative to the full population in years 1979-1985 using the National Longitudinal Mortality Survey (raw numbers reported in sheet Kopzuk-Saez_mortality). Kopczuk and Saez (2004) use this mortality differential for all years. Cols. [3]-[5] and [8]-[10] report the relative mortality for top 10%, top 5%, and top 1% wealth holders for men and women. Wealth is measured at the family tax unit level.