NBER WORKING PAPER SERIES

RECENT TRENDS IN U.S. TOP INCOME SHARES IN TAX RECORD DATA USING MORE COMPREHENSIVE MEASURES OF INCOME INCLUDING ACCRUED CAPITAL GAINS

Jeff Larrimore
Richard V. Burkhauser
Gerald Auten
Philip Armour

Working Paper 23007
http://www.nber.org/papers/w23007

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
December 2016, Revised June 2017

All opinions are those of the authors and should not be attributed to the U.S. Treasury, the Federal Reserve Board, the Federal Reserve Banks, or their staff. We thank Jonathan Fisher, Wojciech Kopczuk, and Maggie Jones, along with session participants at the National Tax Association annual conference, the Association for Public Policy Analysis and Management fall research conference, and the American Economic Association annual conference for their helpful comments on earlier drafts of this paper. Views and opinions expressed are those of the author and do not necessarily represent official positions or policy of the U.S. Department of the Treasury or the National Bureau of Economic Research.

At least one co-author has disclosed a financial relationship of potential relevance for this research. Further information is available online at http://www.nber.org/papers/w23007.ack

Recent Trends in U.S. Top Income Shares in Tax Record Data Using More Comprehensive Measures of Income Including Accrued Capital Gains
Jeff Larrimore, Richard V. Burkhauser, Gerald Auten, and Philip Armour
NBER Working Paper No. 23007
December 2016, Revised June 2017
JEL No. D31,H24,J3

ABSTRACT

Access to IRS personal income tax records improves researchers' ability to track U.S. income and inequality, especially at the very top of the distribution (Piketty and Saez 2003). However, rather than following standard Haig-Simons income definitions, tax form income measures were designed to implement the Internal Revenue Code. Using IRS tax record data since 1989 statistically matched to Survey of Consumer Finances and Census data for income sources not available in tax data, we explore the robustness of levels and trends in inequality using the top income literature's tax return market income definition (Saez 2016) compared to more comprehensive income measures. We find that focusing solely on market income misses the important redistributive effects of government taxes and transfers. In addition, we find that the use of taxable realized capital gains changes the level and trend in top incomes relative to an accrued capital gains measure that is more consistent with Haig-Simons income definitions.

Jeff Larrimore
Federal Reserve Board
20th and Constitution Ave
Washington, DC 20551
jeff.larrimore@frb.gov

Gerald Auten
Office of Tax Analysis
U.S. Department of the Treasury
1500 Pennsylvania Avenue, NW
Washington, D.C. 20220
Gerald.Auten@treasury.gov

Richard V. Burkhauser
Cornell University
Department of Policy Analysis & Management
259 MVR Hall
Ithaca, NY 14853-4401
and University of Melbourne
and also NBER
rvb1@cornell.edu

Philip Armour
RAND Corporation
1776 Main Street
Santa Monica, CA 90401-3208
parmour@rand.org

I. Introduction

An important new international literature (Atkinson, Piketty, and Saez 2011) based on personal income tax return data has focused on the share of income held by top income groups and how it has changed over time. Piketty and Saez (2003) were the first to use Internal Revenue Service (IRS) personal income tax record data to track U.S. levels and trends in income and its distribution in this way. These administrative records offer substantial advantages over survey-based data with respect to their sample size, high response rates, and lower recall bias.

Federal individual income tax rules and forms are intended to implement the Internal Revenue Code, however, and are not necessarily comparable with the income definitions economic researchers prefer to use to measure income.[1] Atkinson, Piketty, and Saez (2011, p. 34) in their review of the results of research based on tax record data state that: "In all cases, the estimates follow the tax law, rather than a 'preferred' definition of income, such as the Haig-Simons comprehensive definition, which includes such items as imputed rent, fringe employer benefits, or accruing capital gains and losses." We add the importance of government to this list by including cash and in-kind government transfers and netting out government taxes.[2]

Unless supplemented with such data from other sources, researchers using IRS tax return data will miss any non-taxable income that does not appear on IRS tax forms. Particularly relevant for research on top income shares, whereas tax record–based researchers sometimes include a measure of taxable realized capital gains as an alternative for accrued capital gains, we show that doing so not only misses capital gains that are not taxed but also fails to reflect the

[1] The U.S. income tax, in its various provisions, can be (or has been) described as a hybrid tax that is a combination income tax, consumption tax, and gross receipts tax.

[2] Including government taxes and transfers to more comprehensive measures of income is standard practice in the survey literature. This method is recommended by the Canberra Group (2011) and the OECD (d'Ercole and Förster 2012) and is also consistent with a Haig-Simon comprehensive definition of income.

year in which these realized capital gains were accrued. As a result, the use of taxable realized capital gains will dramatically alter levels and trends in the share of income flowing to the top 1% in a given year relative to a measure using accrued capital gains.

This paper explores the impact of how income is defined on levels and trends in top income shares. Using income tax records from the IRS Statistics of Income, with a statistical match to Survey of Consumer Finances (SCF) and March Current Population Survey (CPS) data for income sources that cannot be observed in IRS data, we consider the extent to which trends in top 1% income shares differ when using a narrow tax return–based income definition compared to broader income definitions more in the spirit of Haig-Simons income principles.

This research makes several substantial advancements relative to previous research that considered how using more comprehensive income definitions influence inequality trends (see, e.g., Burkhauser, Larrimore, and Simon 2012; Armour, Burkhauser, and Larrimore 2014; Smeeding and Thompson 2011).

First, by starting with income tax data from the IRS rather than survey-based data, we are able to capture the trend in top 1% income shares using our broader income definitions and compare them directly to top 1% income shares from Piketty and Saez (2003) that focus on market income from tax returns.[3] By contrast, most previous research considering the impact of income definitions on inequality trends have exclusively used survey data that is less able to track the top of the income distribution.[4]

Second, we include estimates of accrued gains on housing based on individual-level

[3] Researchers using survey-based data have typically avoided consideration of the top 1% income share because of concerns about the coverage of the survey data at the top of the distribution as well as topcoding of data to protect the confidentiality of high-income respondents. For additional information on these limitations of survey-based data for considering top income shares, see Atkinson, Piketty, and Saez (2011), Burkhauser et al. (2012) and Burkhauser et al. (2016).

[4] Auten and Splinter (2016) show the importance of the base-broadening provisions in the Tax Reform Act of 1986 on income inequality trends before and after 1986.

property values from property-tax records and data on local-level housing appreciation. Doing so provides a more accurate assessment of the capital gains from housing than does previous research, which either ignored capital gains from housing or used only national-level housing estimates (see, e.g., Armour, Burkhauser, and Larrimore 2014; Piketty, Saez, and Zucman 2016). To our knowledge, this is the first paper to consider how capital gains from local-level housing price trends impact broader measures of income inequality such as top income shares.

When we evaluate distributional trends using broader income measures that include our improved estimate of housing capital gains, we observe trends that differ in important ways from those found using only income as it appears on income tax returns. Most notably, while Saez (2016) and others using tax data including taxable realized capital gains find that top income shares fell during the early years of the Great Recession, we find—using our accrued capital gains measure that fully captures the collapse of the housing market in 2008 and its disproportionate negative consequences for the American middle class—that these shares dramatically increased.

II. Defining Income

Recognizing that the choice of income definition may influence income trends, what is the most appropriate way to measure income? The traditional view in the economics literature is that an ideal income definition captures the total inflow of resources that individuals receive for their potential personal consumption in a year, regardless of who provides the income or the form it takes. This principle underlies the Haig-Simons income definition, which states that individuals' yearly income is equal to their consumption plus the change in their net wealth in that year (see Auerbach 1989 and Barthold 1993 for discussions of the Haig-Simons income

approach and Haig 1921 and Simons 1938 for the original sources). On the income side of the Haig-Simons equation, this implies that income should include any consumable resources flowing to individuals in a given year. This approach includes not only before-tax cash income but also in-kind employee benefits, imputed rents from owner occupied housing, and accrued capital gains. It also recognizes the importance of government taxes and transfers by including cash and in-kind government transfers and netting out government taxes.

Despite general agreement that the comprehensive Haig-Simons income measure is the gold standard for defining economic income, by necessity most researchers base their choice of income definition on data availability. For example, some researchers using IRS tax records data limit their analysis to pre-tax, pre-transfer income of tax units since non-taxable sources of income are not included in these data (Piketty and Saez 2003). Similarly, researchers using March CPS data typically include transfer income, but generally exclude in-kind transfers, taxes, and all capital gains—none of which are captured in CPS data (see, e.g., Burkhauser, Feng, Jenkins, and Larrimore 2011; Gottschalk and Danziger 2005).[5]

One of the more important aspects of measuring comprehensive incomes in the spirit of Haig-Simons is the appropriate treatment of capital income and returns to asset wealth. The four major approaches to handling these returns to capital income used in the literature are: 1) ignoring all capital gains (Aguiar and Bils 2016; Burkhauser, Larrimore, Simon 2012; Proctor, Semega, and Kollar 2016); 2) including capital gains at realization as they appear on tax returns (Congressional Budget Office 2016; Piketty and Saez 2003); 3) distributing corporate-retained earnings in lieu of capital gains (Piketty, Saez, and Zucman 2016); and 4) distributing capital

[5] Notably, while the Earned Income Tax Credit is among the largest cash transfer programs for low income individuals, since it is administered through the tax system rather than as an independent transfer program, it is not included in the CPS questionnaire and is excluded from both the Census Bureau's official pre-tax, post-transfer income measures and most research using this income definition.

gains as they accrue (Smeeding, and Thompson 2011; Armour, Burkhauser, and Larrimore 2014).[6]

The first of these approaches—ignoring all capital gains—almost certainly understates incomes, particularly among higher-income individuals and homeowners who may have substantial capital income. Yet it is the dominant way income is measured in the income and poverty survey-based literatures. The second—including capital gains at realization—is convenient, as tax returns offer high-quality data on realized capital gains on taxable assets (see, e.g., Piketty and Saez 2003; Saez 2016). However, as described in Armour, Burkhauser, and Larrimore (2014), many realized capital gains are never reported on tax returns, including most gains from primary housing, those occurring in certain tax deferred accounts, and those on assets held until death. Furthermore, the realized capital gains that do appear on tax returns are often not reported until years or even decades after they were accrued.[7] This factor, in turn, impacts both the observed level of capital gains and the timing of their receipt.

The third approach—distributing retained earnings to shareholders— has been used by researchers attempting to align individual incomes with national accounts (Piketty, Saez, and Zucman 2016). This approach is advantageous for its alignment with some national accounts measures. But since it distributes only current year corporate incomes to individuals, rather than the value of a corporation based on its projected future earnings as reflected in stock prices, this

[6] Although it does not directly address capital gains, a fifth, less common, approach to incorporating income from wealth is to include the imputed annuitized value of wealth holdings with income regardless of whether that wealth generates any income (Wolff and Zacharias 2009; Wolff, Zacharias, and Masterson 2012). This approach is useful for considering the inequality of potential consumption in a year, but it systematically overstates the income of all individuals with a positive savings rate. This overstatement occurs because under this measure saved income is included in both the year it is earned and in all subsequent years until it is spent. Since this approach reflects a hybrid of income and wealth rather than annual income alone, we do not consider it further here.

[7] Additionally, when sales of assets do occur, it is often not for consumption but rather to change investment vehicles. In early research, Feldstein and Yitzhaki (1978) observe that two-thirds of the value of stock sale are "financial switches," whereby the proceeds are reinvested in another stock or financial asset within one year. Individuals who sell one financial asset and re-invest the proceeds in the same type of financial asset are functionally similar to those who simply hold their financial assets without buying or selling. They will appear differently on tax returns, however, since the former will report taxable realized capital gains in the year and the latter will not.

method fails to fully capture capital income that occurs based on investors' perceptions of a corporation's economic potential. As a result it does not necessarily reflect the change in the price for which investors could sell their asset on the open market in any given year. Additionally, simply distributing retained earnings will result in substantially lower levels of observed long-run capital income than is observed in asset prices.[8] The approach also does not incorporate capital income from housing, which is not included with corporate retained earnings.

The final approach—to capture capital gains as they accrue in each year—is most consistent with Haig-Simons income principals and is the method we employ in this paper. Including capital gains at accrual, rather than at realization, is commonly cited as a preferred approach for measuring capital gains (see, e.g., Atkinson, Piketty, and Saez 2011; Roine and Waldenström 2012; Slemrod 2016; Smeeding and Thompson 2011; Veall 2012), although data availability often limits its implementation. Including capital gains at accrual is the approach specified by the System of National Accounts, the international standard for national accounting (European Commission et al. 2008), and capital gains are similarly included in this way in the Integrated Macroeconomic Accounts produced by the Federal Reserve Board and the BEA (Bond, Martin, McIntosh, and Mead 2007).

Although many researchers agree that accrued capital gains is conceptually preferable to realized capital gains for measuring income, a major limitation of this approach is the lack of accrued capital gains data. As a result, researchers using this approach must impute accrued capital gains on each of the assets held in household wealth portfolios. Smeeding and Thompson (2011) do so by assuming that all assets receive the long-run average return for the asset class

[8] This difference can be observed by comparing the accrued capital gains revaluation series in the Bureau of Economic Analysis' (BEA) Integrated Macroeconomic Accounts (Table S.3.a) to the undistributed corporate profits series in the BEA's National Income and Product Accounts (Table 1.12). From 1989 through 2013, the BEA reported $8.2 trillion in corporate retained earnings, compared to $15.9 trillion in accrued capital gains from equities going to households and non-profits serving households, along with another $2.4 trillion in accrued gains from mutual funds going to these groups (Bureau of Economic Analysis 2016a, 2016b).

and Armour, Burkhauser, and Larrimore (2014) assume that all assets receive the current-year return for the asset class. The advantage of the Smeeding and Thompson approach is that it smooths the substantial year-to-year variance in accrued capital gains. Nevertheless, doing so is inconsistent with the Haig-Simons principal since it systematically overstates capital gains in years the actual rates of return are low and understates them in years the actual rates of return are high.

The analysis in the current paper builds on the Armour, Burkhauser, and Larrimore (2014) method, which more closely captures accrued capital gains each year. However, we present a new approach to measuring gains on owner-occupied housing that substantially improves upon prior approaches for imputing housing incomes, as discussed further in our Data section. Because of the regional differences in the housing bubble and crash over the last 15 years, this new approach provides a more accurate and nuanced picture of the pattern of these gains. Throughout this paper, we focus on key income metrics from the inequality literature. We consider two base-income measures and three treatments of capital gains.

The first income measure is *tax return income*, which includes labor earnings and non-labor market income such as small business income, farm income, taxable and tax-exempt interest, dividends, rents, royalties, and taxable and non-taxable Social Security benefits. This is a broader income measure than the *tax return market income* measure used by Piketty and Saez (2003) and Saez (2016), in that it includes the non-market income sources that appear on IRS tax returns: specifically, Social Security benefits and unemployment insurance. It also differs by adding back the foreign earned income exclusion and by deducting gambling losses from gambling winnings for those who itemize (reflecting that net gambling winnings are a more accurate reflection than gross gambling winnings for this form of income). In addition, alimony

paid and state and local tax refunds (which adjust for over-deduction of taxes in the prior year) and net operating losses carried over from prior years are also removed as they do not reflect current year net income. Finally, to reflect that business expenses are part of the cost of generating income rather than pure consumption, we exclude from income the net employee business expenses that appear on tax returns.[9]

The second income measure, *comprehensive income*, includes all elements of *tax return income* but also includes federal income and payroll tax credits or liabilities along with major cash transfers, in-kind transfers, imputed rents from owner occupied housing, and in-kind benefits that do not appear on tax returns. The untaxed cash transfers include workers compensation, supplemental security income, public assistance income, child support income, and other financial and educational assistance as captured in the March CPS. The included in-kind transfers are the ex-ante value of employer- and government-provided health insurance, food stamps, housing subsidies, and school lunches. This measure also includes imputed rents as part of income, which reflects the resource flow from homeownership and is an important component of the Haig-Simons income definition. While Andrews, Sanchez, and Johansson (2011) document that imputed rents are currently taxed in several countries (Iceland, Luxembourg, the Netherlands, Slovenia, and Switzerland) and hence would be included in a *tax return income* measure in those countries, imputed rents are not taxed in the vast majority of countries, including the United States. Some researchers (e.g., Frick, Grabka, Smeeding, and Tsakloglou 2010; Piketty, Saez, and Zucman 2016) estimate and include imputed rents in their

[9] These expenses are observed only to the extent that they exceed 2 percent of AGI income for taxpayers who itemize their deductions. Although all employee business expenses should in theory be netted against income, we cannot observe such expenses for those whose net employee business expenses are less than 2 percent of AGI or for non-filers. As a result, to avoid treating someone whose net employee business expenses are just above the 2 percent threshold and someone whose expenses are just below the threshold (or who does not itemize) differently, we only remove from income the portion of reported employee business expenses that are above the threshold. The expenses we remove from income are thus likely to reflect expenses that are unusually high relative to income, especially in certain occupations.

measures of income, but most researchers do not. Imputed rents are included in the BEA's Gross Domestic Product estimates in the United States.[10]

For each of these income measures, we also evaluate the impact on top income shares of using the three distinct capital gains treatments described above: excluding capital gains completely, including taxable realized capital gains, and including all capital gains at accrual.

III. Differences between Taxable Realized Capital Gains and Accrued Capital Gains

Taxable realized capital gains and accrued capital gains differ in several important respects. Taxable realized capital gains measure the gains at the point an asset is sold, rather than at the point the asset appreciates in value. Thus, if an investor purchases an asset in 1990 for $10,000, which appreciates to $40,000 by the year 2000, but remains largely flat thereafter until it is sold in 2010, the $30,000 gain would appear on the investor's 2010 tax return even though virtually all the investment returns accrued in the 1990s.[11] This method delays the timing of when the gains appear in the data. Furthermore, it can also result in an artificial increase in observed inequality when multiple years of capital gains from an asset are bunched into a single year. IRS Sale of Capital Assets data from 1999 through 2007 show that 97 percent of realized capital gains during this period were on assets held longer than one year, and more than 40 percent were on assets held for over a decade (Table 1). By contrast, accrued capital gains include the change in asset values in the year that the asset appreciates (or depreciates), which

[10] Recognizing that researchers often exclude imputed rents when considering broader income measures, we also compute the comprehensive income series excluding imputed rents. Doing so generally has little impact on the top income shares. The one exception is when including accrued capital gains in 2008. In that year, including or excluding imputed rents substantially affects top income shares, as these rents partially offset the substantial accrued housing losses that occurred in 2008 among homeowners. Results that exclude imputed rents from comprehensive income are available upon request from the authors.

[11] This example follows the investment return of the S&P 500 over this period. $10,000 purchased in 1990 would have been worth $39,991 in 2000 and worth $38,086 in 2010. Consistent with how capital gains appear on tax forms, these values are in nominal dollars and are not adjusted for inflation.

better reflects the timing of gains. This approach may, however, result in an increase or decrease in capital gains in any given year relative to realized gains, depending on the actual rate of return on assets as well as the level of capital gains realizations.

A second difference between the series is that taxable realized capital gains exclude important classes of capital gains. First, this approach excludes all realized capital gains occurring in tax-preferred accounts. In the case of traditional Individual Retirement Account (IRA) returns, these gains are deferred from appearing on tax returns until retirement, when the funds are withdrawn from the account and reported as ordinary income. In the case of Roth-IRA returns, the capital gains never appear on tax returns. It also excludes all realized capital gains on assets which are held until death, at which time the cost-basis of the asset adjusts to the value at death so decedents owe no capital gains on the asset upon the sale (except for gains occurring after the death). Furthermore, taxable realized capital gains exclude most capital gains on housing assets. Current tax laws exclude the first $250,000 of capital gains on one's primary residence ($500,000 for married couples) from tax and from reporting on tax returns. Since the median sales price on existing homes sold in August 2016 was $240,200 (National Association of Realtors 2016), the vast majority of capital gains on housing are excluded from tax returns.[12] In addition, taxpayers can exclude 50 percent (100 percent of new investment starting in 2010) of up to $10 million of qualified business stock gains held at least five years and meeting various requirements. The exclusions for these asset classes lower the observed levels of capital gains captured as realized taxable capital gains in the tax record data, although the precise impact on

[12] Auten and Gravelle (2009) found that only about 350,000 or 6.5 percent of the 5.7 million sales of existing homes were reported on tax returns. Complicating long-term trends in taxable realized capital gains from housing is the changes to housing capital gains treatment from the Tax Reform Act of 1997. Prior to 1997, housing capital gains were subject to taxation if the seller did not purchase a new home of equal or greater value within two years, although because of the rollover of gains into new homes relatively few gains on housing were taxed. An additional complication was a once-per-lifetime housing capital gains exclusion of $125,000 for individuals who were age 55 or older. Cunningham and Engelhardt (2008), Shan (2008), and Auten and Gravelle (2009) describe these changes and discuss their impact on homeowner mobility.

the measure of the top 1% income shares depends on where in the distribution these non-observed assets are held.

A third limitation of taxable realized capital gains is that the full value of net capital gains are included in Adjusted Gross Income (with the exception of those on certain tax-preferred assets), whereas net capital losses are excluded to the extent that they exceed $3,000.[13] Capital losses exceeding the $3,000 limit may be carried forward to offset future capital gains but may not offset ordinary income. Capital losses on principal residences are never observed because capital losses on personal assets are not deductible. This factor was particularly important during the Great Recession and its aftermath, when middle-class families who sold or experienced a foreclosure on their houses may have realized substantial capital losses.[14] The taxable realized capital gains series will miss this aspect of the recession.

Finally, a fourth limitation of taxable realized capital gains is that realizations are sensitive to the capital gains tax rate, which influences the willingness of investors to sell appreciated assets and realize the gain (e.g., Dowd, McClelland, and Muthitacharoen 2012 estimate a long-run tax elasticity of realized capital gains of –0.79). This condition was particularly important in the 2000s, when the capital gains marginal tax rate in the Economic Growth and Tax Relief Reconciliation Act (EGTRRA) and the Jobs and Growth Tax Relief Reconciliation Act (JGTRRA) tax reforms in 2001 and 2004 lowered the maximum tax rate for long-term capital gains from 20.17 percent in 2001 to 15.7 percent in 2007. It can also be observed in 2013, when the large increase in the top capital gains rate (from 15 percent to 23.8 percent) caused an acceleration of realized gains from 2013 into 2012 (US Department of

[13] Realized capital losses in excess of the loss limit are fully reported on Schedule D, but are not carried to the 1040 form.

[14] A portion of this debt may be observable in tax data using the 1099-C, which is the debt written off by lenders in foreclosure. However, this would capture only a small subset of all capital losses and to our knowledge no researchers have attempted to incorporate these losses into realized capital gains calculations.

Treasury 2016). As a result, observed realized capital gains, and the top 1% income shares, are sensitive to the behavioral responses of capital gains realizations to tax policies.[15]

Accrued capital gains have neither missing asset class concerns, missing capital loss concerns, nor concerns regarding behavioral responses of realizations to tax policies. When considering accrued capital gains, gains or losses on all assets are included, although at the time the asset appreciates in value rather than when the asset is sold. Because assets fluctuate in value from year to year, however, accrued capital gains do exhibit greater volatility than is seen for taxable realized capital gains. As such, top income shares that include accrued capital gains also exhibit higher levels of volatility.

IV. Data

The primary data in this paper are from the Individual Tax Files from the IRS Statistics of Income division. These data are used directly for capturing tax return income (including wages, taxable and tax-exempt interest, dividends, self-employment and small business income, pension and retirement income, unemployment income, and Social Security income), taxable realized capital gains, and federal income and payroll tax liabilities of tax filers.

Although administrative tax return data is valuable for its high sampling rates at the top of the income distribution, they do not observe non-filers or capture all the income sources of those who do file. Adjustments and additions to the base tax data are made to address these concerns as follows. (See the Data Appendix for a fuller discussion of each of the subsections below.)

[15] Realized long-term capital gains also increased after 1993 when the top rate on ordinary income increased and in 1997 when the top capital gains tax rate was reduced from 28 to 20 percent. While occurring before the starting point of our analysis, substantial changes to the capital gains tax rate in the 1986 Tax Reform Act also influenced the timing of capital gains realizations. Auten, Splinter, and Nelson (2016) discuss taxpayer reactions to these changes in tax legislation, including shifting of income to minimize tax liabilities.

A. Non-filers and Dependent Filers

Recognizing that not all individuals file a tax return, we incorporate non-filers by following the approach of Piketty and Saez (2003). They, and we, estimate the total number of potential U.S. tax units each year based on Census Bureau survey data.[16] We also follow their assumption that the market income of non-filers is 20 percent of the mean market income of filers in each year.[17] Although this approach does not yield an actual distribution of non-filers in the population, more than 80 percent of non-filers have income under $50,000 (Cilke 2014) and it is generally believed that virtually all non-filers are well below the 90th percentile of income. Therefore, for measuring top income shares, assuming that non-filers are below the 90th percentile threshold, rather than knowing their full distributional characteristics, is sufficient to identify the number of non-filers and the total income they receive.

Although we generally follow Piketty and Saez (2003) in our treatment of non-filers, we diverge with respect to non-resident filers, dependent filers, and other filers who are under age 20. Piketty and Saez include these tax returns in the same way as all other tax filers, although based on the potential tax unit definition they are not included as potential tax units. This creates an unusual result in 2007, in which (as a result of increased filing for stimulus payments), there are more tax units who file a return than there are potential tax units. To correct for this concern, we drop non-resident filers and dependent filers who are under age 20 from the sample prior to calculating results and adjust the estimated number of tax units by the number of dependent filers

[16] Under our approach, each single individual or married couple age 20 or older in the United States represents one potential tax unit, which roughly matches the definition from Piketty and Saez (2003) and Burkhauser et al. (2012). Tax units represent individuals who file a tax return together, and potential tax units represent individuals who either file together or would be expected to file together if they file a return. This definition should not be confused with households or families, which are common units of aggregation in survey-based research (see, e.g., Burkhauser, Larrimore, and Simon 2012; Bricker et al. 2016; and Larrimore, Mortenson, and Splinter 2017 for discussions of these differences and their impacts on distributional statistics).

[17] Although we follow the non-filer imputation of Piketty and Saez (2003), Auten and Splinter (2016) observe that this approach may understate the true income of non-filers and that their income levels are closer to 30 percent of the mean income of filers in each year. Increasing the imputed income for non-filers would decrease the observed top income shares in all years.

age 20 and over (primarily full-time college students).

B. Cash and In-kind Transfers not Reported on Tax Returns

To incorporate income beyond that which appears on tax returns, we integrate data from the IRS, other administrative agencies, the Census Bureau's March CPS, and the Federal Reserve Board's Survey of Consumer Finances.

Social Security benefits of filers are captured directly on tax forms (including Form 1099-SSA). To incorporate the benefits of non-filers, we compare the total Social Security payment outlays reported by the Social Security Administration to the total reported on tax forms by filers, and assign the residual to the non-filing population.

Cash and in-kind transfers that are not reported on tax forms are added to the dataset using a statistical match to CPS data. Since tax units are not delineated in the CPS data, we divide households into tax units following the guidelines from Piketty and Saez (2003) to estimate potential tax units described above and in our Data Appendix.[18] For each tax unit in the CPS data, we then determine their centile in the taxable income distribution, average their in-kind transfer income from each source for that centile of the distribution, and assign that in-kind transfer income to tax units in the IRS data in the same centile. Recognizing that major transfers, including health insurance provided through Medicare, are correlated with the respondent's age, when imputing the value of cash and in-kind transfers from the CPS data for each tax unit, we split the sample into those under age 65 and those over age 65 prior to computing income centiles (based on the oldest member of the tax unit). Hence, for matching to the CPS data each individual's in-kind incomes are taken from the CPS data based on their centile within the income distribution of those in the same age class. As a result, tax units with people who are over

[18] Similar guidelines have previously been used to estimate tax units in the CPS data by Armour, Burkhauser, and Larrimore (2014), Burkhauser, Larrimore, and Simon (2012), and Burkhauser et al. (2012).

age 65 have substantially higher estimated Medicare benefits than those containing only younger individuals. When incorporating these in-kind incomes, we follow the approach of Burkhauser, Larrimore, and Simon (2012), the Congressional Budget Office (2016), and Piketty, Saez, and Zucman (2016) and assign tax units the full ex-ante value of employer- and government-provided health insurance.[19]

We incorporate imputed rents by distributing the total imputed rental income of primary residences as reported in the BEA's Gross Domestic Product calculations to each homeowner in the data based on their estimated house value. Our approach for estimating home values for each taxpayer is based on the individual's property tax payments reported on their tax form and the ratio of home values to property tax payments in their local area, as described in the subsequent section outlining the calculation of accrued capital gains on housing.

The relative magnitudes of each of these components of comprehensive income can be observed in Appendix Table A2, which shows the aggregate income in our data from each broad category of income in the comprehensive income measure.

C. Accrued Capital Gains

Perhaps the most important additions to the tax data are the estimates of accrued capital gains for each tax unit, including the accrued gains from housing, which is the primary financial asset for many families. To estimate accrued gains, we build on the methods of Armour, Burkhauser, and Larrimore (2014), who estimated accrued capital gains as the product of the

[19] Based on the observation of Finkelstein, Hendren, and Luttmer (2015) that only 20 to 40 percent of the welfare benefits of Medicaid accrue to beneficiaries, some have argued that a more appropriate method for valuing health insurance benefits is to value them at 20 to 40 percent of their ex-ante value. Since Finkelstein, Hendren, and Luttmer's (2015) estimate comes in large part due to the uncompensated care that uninsured individuals receive, were an adjustment for this care to be included, we believe it would be more appropriate to add the value of this uncompensated care to those without insurance than to subtract its value from those with Medicaid or Medicare. Were we to add a value of uncompensated care for the uninsured to our income definition, it would reduce the top 1% income share further relative to that reported here. Were we to value Medicaid and Medicare at only 20 to 40 percent of the ex-ante value, it would increase the top 1% income share from that reported here.

underlying value of each asset at the start of a period and the rate of return from that asset class. This approach is based on the gross capitalization technique for relating wealth to the income it generates, which has also been used by Piketty, Saez, and Zucman (2016), Saez and Zucman (2016), and Smeeding and Thompson (2011).[20]

In estimating annual accrued gains for owner occupied housing, it is important to account for the diverse experiences of different regions and local areas with respect to the trends and timing of changes in home prices. Since home equity accounts for a large part of wealth holdings for many tax units, tracking changes in accrued capital gains and losses from housing is potentially important in measuring income over the time period of our data. In this paper, for what to our knowledge is the first time, we track these changes by taking advantage of the property tax deduction information available in the income tax data.

To calculate housing capital gains, we start by estimating the market value of houses among homeowners. We first estimate the ratio of mean home market value to mean property tax payment in each county using the county-level home values and property tax from the 1990 and 2000 Decennial Census and from the American Community Survey since 2005. (The property tax ratios since 2005 using the ACS are from Harris and Moore 2013.)[21] This estimate may differ from the statutory property tax rates in local areas to the extent that the assessed values of homes for property tax purposes do not match current market values as reported in the surveys.

For any tax units who file a tax return and itemize their deductions—which comprises the

[20] One concern with this approach is that those with higher incomes may be more knowledgeable or skilled at investing and therefore receive a higher rate of return. However, Saez and Zucman (2016) provide evidence that high-income individuals do not, in fact, receive higher rates of return on their assets than those farther down in the distribution. Lundberg and Waldenström (Forthcoming) consider the gross capitalization model in Sweden and observe that the gross capitalization model accurately reflects the relationship between dividends and corporate stock value for approximately two-thirds of cases, although they do not explore the relationship between asset values and overall investment returns.

[21] Recognizing that the market value of homes to property taxes paid within a county is relatively flat over time except for when the property tax rates are changed by state or local legislators, we use a weighted average of the closest observed years to determine this ratio in the intermediate years between the 1990 and 2000 Decennial Census and between the 2000 Decennial Census and the start of the American Community Survey in 2005. This procedure is modified for two states that had major school finance reforms that resulted in substantial decreases in the mid-decade. In these cases, the ratio from the two bounding decennial census are used on each side of the reform.

vast majority of high-income taxpayers—we then multiply the property tax payments that they report on their tax return by the home-value to property-tax ratio for their county to obtain an estimate of the market value of their specific home. To further improve coverage for taxpayers who remained in the same zip code over three or four years, and for whom property tax information is missing for the middle year, the missing year of property tax data is interpolated based on the surrounding years. Since this approach uses information derived from the specific valuation of their house (property tax payments) along with the local level information on how these values relate to properties in their area of residence, it provides a more detailed assessment of home values than can be obtained using national level information.

Although this approach provides improved estimates of home values for itemizers, it does not pick up the universe of homeowners since not all homeowners file a tax return and not all tax filers itemize their deductions. However, nearly all taxpayers in the top centiles itemize their deductions (96 percent of the top 1% in 2007), so the individuals for whom property-tax records are missing are almost exclusively outside of the top centiles of the distribution. To capture the market value of owner occupied homes for non-filers and non-itemizers, we observe the difference between aggregate home values captured among itemizers using the approach above and the aggregate home values reported by the BEA's Integrated Macroeconomic Accounts (IMA). This residual home value is then distributed among non-itemizing tax units.[22]

Having estimated the current market value of houses owned by each tax unit, we then calculate the accrued capital gains from housing for itemizers by multiplying the home value by

[22] The precise distribution among the non-itemizers will not impact results unless the home values in this group are sufficiently concentrated to push individuals into, or out of, the top 1% of the distribution.

the home price appreciation rate from CoreLogic for their county of residence.[23] For non-itemizers and non-filers, because we do not know which specific tax units are the homeowners to use local-level price trends, we use the national-level home price appreciation rate from CoreLogic. This, in turn, provides our best estimate of the capital gains from housing in each year among those who own a home.

For accrued capital gains from publicly traded and private business investments, the approach is similar. Following the procedure used by Armour, Burkhauser, and Larrimore (2014), accrued gains on investments build on the wealth data from the Survey of Consumer Finances (SCF).

A strong predictor of whether individuals have any business investments, which would generate capital gains, is whether they have any dividends or rents/business income. Each tax unit is grouped based on whether they have neither of these types of income, dividends only (over $100), rents/business income only (over $500 in absolute value), or both. Within each group, individuals are then arrayed into centiles of their total income. For each centile within each of the four groups, the publicly and privately traded business assets are estimated to be the average holdings of individuals in the SCF data in the same centile of the group-level income distribution.[24]

For each asset class, we then impute the rate of return based on the average return of that asset class in the given year. For stocks, we do so using the rate of return of the S&P 500 in the specified year. For private business wealth, we use the implied rate of return on non-corporate

[23] The CoreLogic Home Price Index tracks home prices using repeat sales transactions. Not all counties have sufficient sales volume for CoreLogic to compute a county-level home price index. In these cases, the state level appreciation rate is used. For additional information on the CoreLogic index, see http://www.corelogic.com/products/corelogic-hpi.aspx

[24] The SCF organizes individuals into Primary Economic Units (PEU), which may consist of one or more tax units. As a result, in 2013 there are approximately 122 million PEUs, compared to 162 million tax units. Prior to computing the average asset holdings of those at each point in the distribution, it is necessary to split PEUs into tax units—which we do evenly throughout the income distribution.

business equity from the revaluations series in the IMA.[25]

Because the SCF is a triennial survey, to produce annual results the distribution of assets in non-survey years is estimated using a weighted average of the previous and subsequent survey. The rate of return for these intermediate years is still the actual return from the specified year from the S&P 500 and IMA data.

V. Results

A. Trends in Tax Return Income with and without Capital Gains

In Figure 1, we start by comparing the top 1% income shares of tax units for our tax return income series to the top 1% tax return market income shares that are observed by Piketty and Saez (2003), both with and without taxable realized capital gains. The inclusion of all income sources that appear on tax returns, including Social Security and Unemployment Insurance as well as the other refinements described above, results in top income shares that are between 1 and 2 percentage points lower in most years—both when excluding and including taxable realized capital gains.[26]

Considering the top income share trends for these baseline series, the top 1% income share for tax return income increased over the two-business cycles prior to the start of the Great

[25] The IMA provides the revaluations of several asset classes, including nonfinancial noncorporate equity. These revaluations are similar in concept to our annual accrued capital gains. To determine the implied rate of return, we take the ratio of revaluations from nonfinancial non-corporate equity in each year to the prior-year end of year total asset value from nonfinancial noncoporate equity. This approach is a divergence from the rate of return for private business wealth assumed by Armour, Burkhauser, and Larrimore (2014) who used the S&P 500 for private wealth as well. Because the IMA-implied return on non-corporate business equity was lower in the late 1980s than S&P returns for publicly traded businesses, using the IMA return lowers the top income shares using this approach at the early years in our period and makes top income share growth more positive. Results using the S&P 500 rate of return for private business wealth instead would show slower top income share growth.

[26] When we use only the sources of income from the tax records that Piketty and Saez (2003) use in their tax return market income series, we are able to nearly perfectly replicate their top 1% income shares series shown in Figure 1, so this difference can be fully attributable to the fuller income measure and corrections to the tax record sample, rather than other methodological variations.

Recession with or without the inclusion of taxable realized capital gains.[27] However, as was observed previously by Piketty and Saez, the level of top incomes and its growth is accentuated when taxable realized capital gains are included in the calculation. From the business cycle peak in 1989 to the peak in 2007 before the Great Recession, the top 1% share of tax return income without capital gains rose by 5.7 percentage points. By contrast, when including taxable realized capital gains, the top 1% share of tax return income rose by 9.1 percentage points. Thus, reflecting the importance of capital gains, more than one-third (3.4 percentage points) of this 9.1 percentage point increase in top 1% income shares can be attributed to increases in the realized capital gains of those at the top of the income distribution.

During the Great Recession in 2008 and 2009, the top 1% income shares for these series fell – particularly when including taxable realized capital gains. But the trend then reversed and top 1% income shares rose through 2012 before falling again in 2013, when capital gains tax rates were increased.[28]

As discussed above, however, when including capital gains in income calculations, focusing on taxable realized capital gains fails to reflect the true level and timing of the economic income flow. Therefore, Figure 2 reveals how our top 1% income shares series differ if we incorporate gains in the year they are accrued rather than in the year they are realized and reported on tax returns. To do so, we once again show our tax record income series with no capital gains and with taxable realized capital gains from Figure 1. To better show how housing

[27] Throughout the paper, for the analysis of the period before the Great Recession we focus on the 1989 and 2007 business cycle peaks. We do so, in part, because 1989 is the first year for which SCF data is available. However, since 1989 and 2007 are both peaks of business cycles, this comparison is also advantageous as it allows us to consider long-run trends in top income shares while avoiding conflating the comparison with business cycle effects.

[28] Ahead of the 2013 increase in both ordinary and capital gains top rates and the new 3.8 percent tax on net investment income, the highest-income taxpayers—who were the only ones effected—shifted ordinary income and accelerated the realization of capital gains into 2012 (Auten, Splinter, and Nelson 2016; Saez 2016). This shift produced a spike in 2012 in the top 1% income share based on ordinary income or including realized capital gains. This result can be seen in Appendix Table A1, as realized capital gains increased by 60 percent in 2012 and then decreased by over 20 percent in 2013. By contrast, stock prices increased by 30 percent in 2013 compared to only 13 percent in 2012. In 2014, Saez (2016) reports that the top 1% share increased, although not back to the level observed in 2012.

versus investment gains impact top 1% income share trends, we add these accrued capital gains in two steps. We first add accrued capital gains from investments in publicly traded corporations and privately held businesses to our no capital gains series (dashed line with diamonds) and then add the accrued capital gains from housing in a subsequent step (solid line with diamonds).

Considering first the longer-term trend over the two full business cycles before the Great Recession (1989 to 2007), the growth in top 1% income shares using capital gains in the year they are accrued is slower than the growth found using taxable capital gains in the year they are realized. This is the case irrespective of how inclusively we measure accrued gains. The top 1% income share including taxable realized capital gains rose by 9.1 percentage points from 1989 to 2007. Using our most inclusive accrued gains measure—which includes accrued gains from public investments, private business, and housing—the increase in the top 1% income share was approximately 40 percent smaller, rising by 5.3 percentage points from 1989 to 2007.

Both these longer-term trends, as well as their short-run fluctuations, are partially driven by the strength of the equities markets at any given time that result in increases or decreases in capital gains accruals. For example, the increase in the top 1% income share in 1995 corresponds with a 31 percent increase in the S&P 500 in that year, which was followed by continued strength in the equities market for much of the late 1990s (see Appendix Table A1). Similarly, the increase in top income shares in 2003 through 2005 resulted from strength in the rate of return on public equities in 2003 and on privately held businesses in 2004 and 2005. Overall, the average nominal return on the S&P 500 was 13 percent in the 1980s and 16 percent in the 1990s. It was only 1 percent from 2000 through 2007. This decline in the rate of return on public equities reduces the growth in top 1% income shares over the period coming from accrued capital gains when it is not offset by gains in other assets held by the top of the distribution, such

as privately held businesses (as was seen in 2004-2005). After a 38.5 percent decline in 2008, the equities market has been quite strong since 2009, with an average annual return on the S&P 500 of nearly 16 percent per year through 2013, which explains the more recent increases in top 1% income shares.

In addition to more accurately reflecting the timing of when capital gains are received, this series also includes capital gains from housing, which are almost completely absent from the taxable realized capital gains series. Since housing is the main asset of many tax units outside of the top income centiles, incorporating these gains into the income definition reduced the level of top income shares when housing values were increasing in the early 2000s (this can be observed by comparing the series with only capital gains from public and private investments to the most inclusive series that includes housing capital gains as well). But, even in 2007, when housing values began to decline, the top 1% income share is below that observed when only taxable realized capital gains are counted and its increase over the two business cycles since 1989 is slower than that observed when using taxable realized gains.

However, although incorporating accrued capital gains from housing mitigates the growth in top income shares prior to the Great Recession, using this more inclusive capital gains definition results in an enormous spike in top income shares during the first year of the Great Recession and accentuates the growth in top income shares during the recovery. The decline in top 1% income shares we observed for the taxable realized capital gains series in 2008 came primarily from the decline in taxable realized gains among those at the top of the distribution. But it missed the scope of capital losses throughout the distribution, including the capital losses from housing. When incorporating capital gains and losses into income definitions, these substantial losses are an important part of the trend during the recession, and more fully reflect

the magnitude of the recession than do the decline in capital gains realizations. As seen in Appendix Table A3, total accrued capital losses (including housing) in 2008 amounted to $8.2 trillion (compared to tax return income excluding capital gains of $8.6 trillion in Appendix Table A2). Because of the unprecedented scale of the capital losses in 2008 that almost completely counterbalanced income from other sources reported on tax returns in that year, once we include accrued losses we observe that this top 1% income share series exhibited an enormous spike in 2008. In particular, the top 1% share of tax return income including our total measure of accrued capital gains was over 100 percent in 2008. In other words, when focusing on income from tax returns only—but including capital gains and losses in a way that more closely reflects its level and timing—the net income of many taxpayers was actually negative in 2008 because their accrued capital losses were greater than their tax return income from other sources. Although this top 1% income share measure exhibited a dramatic spike in that year, we emphasize that this does not mean that very-high-income individuals were exempt from these capital losses or that their incomes increased during this period. Since centiles of the distribution are always based on the income measure evaluated, the observed spike also reflects a reshuffling of who is in the top 1% of the distribution relative to a series with no capital gains, as some individuals who have high tax return incomes experienced substantial capital losses that caused them to drop out of this top 1% series that captures these losses in its income measure.[29]

The spike in top income shares in 2008 proved to be a single-year event for this series, as top 1% income shares dropped back to 16.5 percent in 2009. But in subsequent years of the economic recovery, stock prices recovered more rapidly than did house prices. As a result, from 2009

[29] This issue reflects a broader challenge with measuring income shares for a single year. Auten, Gee, and Turner (2013) for example, observe that in any given year only about two-thirds of those in the top 1% of the income distribution on tax returns are also in the top 1% in the following year. Since Auten, Gee, and Turner (2013) focus on taxable realized capital gains rather than accrued capital gains, the rates of persistence that they observe will likely differ from those using our measures. In 2008, in particular, persistence of the top 1% from the prior year and subsequent year are likely lower using our accrued gains series due to the magnitude of the accrued losses.

through 2012 the assets of those at the top of the distribution in this series (as well as others with significant investment and business assets) recovered faster than did the investments of those in the middle of the distribution (housing assets). Hence, by 2012 the top 1% income share including these three sources of accrued capital gains increased back to its 2007 level, before falling slightly (0.2 percentage points) in 2013. This result is in sharp contrast to the top 1% income share series using taxable realized capital gains. It fell by 1.2 percentage points in 2013 due to artificial timing shifts in realizations to avoid the tax increases and the 3.8 percent net investment income tax that began in 2013 (Auten, Splinter, and Nelson 2016; Saez 2016). The contrast between 2012 and 2013 observations in these two series nicely shows the extent to which accrued capital gains series are less susceptible to behavioral changes by individuals deferring or locking in gains based on changes in the tax code.

B. Trends in Comprehensive Income with and without Capital Gains

Although tax return income is useful for understanding the distribution of income as captured by the tax code—including income from labor market and investment activities—it presents an incomplete picture of the full distribution of resources to Americans beyond solely its incomplete treatment of capital gains. This is because cash and in-kind transfers that do not appear on tax returns, along with in-kind employer benefits and tax credits and liabilities, are important components of resources for many tax units even though they do not show up in income as observed on tax returns. This approach also misses the imputed rental income of homeowners, which helps accentuate the gains from homeownership in most years, but partially offsets the losses among homeowners in 2008. In addition, tax credits including the EITC, the Child Tax Credit, the 2008 stimulus rebate, the 2009 and 2010 Making Work Pay Credit and homebuyer tax credit, the 2011 and 2012 payroll tax holiday, and the American Opportunity Tax

Credit all have material impacts on the financial resources of their recipients. The spike in top 1% income shares that is observed in 2008 when using only income that appears on tax returns, along with accrued capital gains including housing, partially reflects the fact that the income base from which top income shares are computed fails to include these additional resources that individuals receive.

Figure 3 illustrates the impact of including taxes, transfers, and in-kind benefits on the top 1% income share, excluding capital gains. Using a more inclusive income definition lowers the top 1% income share and slows its growth. Reassuringly, progressive taxes and transfers mitigate income inequality to some degree. The top 1% share in 2013 was 17.4 percent based on the Piketty-Saez narrower tax return market income definition (excluding capital gains). As observed previously in Figure 1, including other pre-tax cash income that is captured on the tax return lowers this share to 15.9 percent. When broadening the income definition even further by including taxes, in-kind transfers, and non-taxable cash transfers, it lowers the top 1% income share in 2013 to 11.1 percent. Thus, even before evaluating the impact of capital gains, using a more comprehensive income measure reduces the top 1% income share in 2013 by 6.4 percentage points relative to the Piketty-Saez income measure and by 4.9 percentage points relative to the tax return income measure.

Just as important as its impact on levels is the impact that including additional income sources has had on the trend of top incomes. Using only tax return income (excluding capital gains), the top 1% income share increases by 5.7 percentage points between over the two business cycle period between 1989 and 2007. Using our comprehensive income definition, the growth over the two business cycles prior to the Great Recession is slower than that for tax return income (and slower still than the growth in top income shares for market income seen in

Figure 1)—increasing by only 3.1 percentage points from 1989 through 2007. Top income shares over the period from 2007 through 2013 fell by 1.6 percentage points from 2007 through 2013 using the more comprehensive income measure.

Though top 1% income shares are now lower, the impact of each of the potential capital gains treatments on our comprehensive income series is substantively similar to that observed previously when we added capital gains to tax return income—with the exception of 2008. Adding taxable realized capital gains to comprehensive income (see Figure 4) increases the level of inequality as well as its growth over time. Adding capital gains at accrual also increases the level of top 1% income shares in most years relative to the no-capital-gains series. But, using accrued capital gains rather than taxable realized capital gains results in a slower growth from 1989 through 2007 in the top 1% income share. Using our preferred income measure—comprehensive income with accrued capital gains from all sources including housing—the top 1% income share rose by only 1.7 percentage points from 1989 through 2007.

While observed inequality growth is slower in the two business cycles prior to the Great Recession when using accrued gains rather than realized taxable gains, we again see that at the start of the Great Recession the top 1% income share increases in 2008 due, in part, to our ability to better capture capital losses, including those from housing, with this measure. Once again, this reflects the fact that when measured at accrual the capital losses for many taxpayers in 2008 exceeded income from other sources. By using a broader income measure that is closer to a full Haig-Simons definition, however, the accrued losses at the start of the recession are less substantial as a share of income from other sources. As a result, although the top 1% income share still rises in 2008, it does so less dramatically than previously shown using the narrower income definition. The top 1% income share in this year was 21.5 percent.

Finally, looking at more recent years, we observe a steady growth in the top 1% share of our most comprehensive income measure during the economic recovery since 2009. The 16.5 percent top 1% income share using this series in 2013 is the highest of any year since 1989, other than in 2008. Despite the increase in the share of comprehensive income going to the top 1%, this 2013 level is still below the 22.8 percent share of market income plus realized capital gains observed by Piketty and Saez in 2012 and the 20.0 percent share that they observe in 2013, as reported in Figure 1. Hence, when considering the full 24-year period for which we have data, using a more comprehensive income measure results in lower levels of top income inequality in most years and slower inequality growth than was previously believed from 1989 through 2007, but more rapid inequality growth from 2007 through 2013.

VI. Discussion and Conclusion

Empirical researchers are ultimately limited by available data in their ability to provide answers to policy questions. Piketty and Saez (2003) in their seminal effort to capture long-term trends in U.S. income inequality were the first to show the value of using tax record data for this purpose. But to do so consistently over the entire period, they had to focus on the share of market income held by top income groups and, to the extent that they include capital gains, had to do so at the point of realization when the gains appeared on tax returns. Using these data, Piketty and Saez observed that the share of market income (both including and excluding capital gains) of the top 1% rose over the first part of the 20th century and fell almost continuously up to the mid-1950s. But more importantly, they found that thereafter this decline levelled off and in the early 1980s began to increase. Especially when including realized capital gains, the increase that Piketty and Saez observe since 1989 has been dramatic—with the top 1% income share

increasing by 61 percent (from 14.5 percent to 23.5 percent) between the two business cycle peak years of 1989 and 2007. The share then fell during the first two years of the Great Recession and rose and fell thereafter through 2013.

Here, with access to much richer contemporary data for the years 1989 through 2013, we show the sensitivity of Piketty and Saez's finding of rapid increases in income equality between 1989 and 2007 as well as the pattern of change since then to the nature of their income definition. We show that moving from a market income to a taxable income to a comprehensive income measure that is more closely aligned with the Haig-Simons definition not only lowers top income shares but also reduces inequality growth up through the start of the Great Recession. But even more importantly, we show that the decision by Piketty and Saez to focus on taxable realized capital gains rather than accrued capital gains greatly affects their patterns of levels and trends not only relative to income measures that do not include capital gains but to one that uses a Haig-Simons preferred accrued capital gains measure.

Figure 5 most directly focuses on Piketty and Saez's preferred measure—market income including taxable realized capital gains—and our two proposed alternatives—comprehensive income with and without accrued capital gains.

Our comprehensive measure without capital gains is similar to their measure in direction but with both lower levels of inequality and muted trends throughout the entire period from 1989 through 2013. At the business cycle peak in 2007 before the start of the Great Recession, the top 1% income share we find using our comprehensive measure without capital gains reached a peak of 12.7 percent, well below the 23.5 percent peak Piketty and Saez report. Additionally, although both top 1% income shares were greater than they were in 1989, our comprehensive measure without capital gains rose less rapidly. Relative to 1989 levels, our top 1% income share

increased by 2.1 percentage points (19.8 percent), compared to the 9.1 percentage point (61 percent) growth they report. The six years since the start of the Great Recession in both series show initial declines in levels and trends in top 1% income share that are only partially offset by increases in later years.

When we compare our comprehensive income including accrued capital gains measure with the Piketty and Saez measure of market income with taxable realized capital gains, our measure is again below their measure in all years between 1989 and 2007 and growth is considerably lower in our series with accrued capital gains compared to their series. But unlike our comprehensive measure without capital gains, the year trends in our accrued capital gains series are quite different and do not peak in business cycle peak years as do the other two series. During each expansion, the growth in top 1% income shares in our accrued capital gains series seem to precede that observed in the other two series.

The contrast between our comprehensive income series with accrued capital gains and the Piketty and Saez series with taxable realized capital gains is also clearly apparent during the Great Recession and its aftermath. During the initial year of the Great Recession, using accrued gains more fully encapsulates the magnitude of the recession, as the capital losses from equities, private businesses, and housing in that year more than offset income from other sources for many tax units. This observation results in a notable increase in the top 1% income share in that year that is not observed in the taxable realized capital gains series. Additionally, during the economic recovery from 2009 through 2013—which exhibited particular strength in the equities markets—the top 1% share increased by more than previously observed using a taxable realized capital gains series (particularly in 2013, where the taxable realized gains series is especially sensitive to the changes in the tax law).

The divergences that we highlight in Figure 5 suggest that in the unsettled question of the place of capital gains in the income literature, it is not sufficient to simply focus on the question of *whether* to include capital gains but, if so, *how* to include them. Only by including capital gains as they accrue, and incorporating capital gains and losses that do not make it onto tax records, are we able to fully observe the experience of tax units within and outside the top 1% of the income distribution during the Great Recession.

References

Aguiar, M., & Bils, M. 2016. Has Consumption Inequality Mirrored Income Inequality? *American Economic Review* 105(9), 2725–56.

Andrews, D., Caldera Sánchez, A., & Å. Johansson. 2011. Housing Markets and Structural Policies in OECD Countries. *OECD Economics Department Working Paper* 836.

Atkinson, A. B., Piketty, T., & Saez, E. 2011. Top Incomes in the Long Run of History. *Journal of Economic Literature* 49(1), 3–71.

Armour, P., Burkhauser, R. V., & Larrimore, J. 2014. Levels and Trends in United States Income and Its Distribution: A Crosswalk from Market Income Towards a Comprehenesive Haig-Simons Income Approach. *Southern Economic Journal* 81(2), 271–93.

Auerbach, A. J. 1989. Capital Gains Taxation and Tax Reform. *National Tax Journal* 42(3), 391–401.

Auten, G., & Gravelle, J. 2009. The Exclusion of Capital Gains on the Sale of Principal Residences: Policy Options. Proceedings of the 102nd Annual Conference of the National Tax Association. Washington DC: National Tax Association, 103–12.

Auten, G., Gee, G., & Turner, N. 2013. Income Inequality, Mobility, and Turnover at the Top in the US, 1987-2010. *American Economic Review 103*(3), 168–72.

Auten, G. & Splinter, D. 2016. Using Tax Data to Measure Long-Term Trends in U.S. Income Inequality. *Working Paper.*

Auten, G., Splinter, D., & Nelson, S. 2016. Reactions of High-Income Taxpayers to Major Tax Legislation. *National Tax Journal* 69(4), 935–64.

Barthold, T. 1993. How Should We Measure Distribution? *National Tax Journal* 46(3), 291–99.

Bond, C. A., Martin, T., McIntosh, S. H., & Mead, C. I. 2007. Integrated Macroeconomic Accounts for the United States. *Survey of Current Business* 87(2), 14–31.

Bricker, J., Henriques, A. Krimmel, J., & Sabelhaus, J. 2016. Measuring Income and Wealth at the Top Using Administrative and Survey Data. *Brookings Papers on Economic Activity*, 261–312.

Bureau of Economic Analysis. 2016a. *National Income and Product Account Tables*. Available online via: https://www.bea.gov//national/nipaweb/DownSS2.asp.

Bureau of Economic Analysis. 2016b. *Integrated Macroeconomic Accounts for the United States*. Available online via: https://www.bea.gov/national/nipaweb/Ni_FedBeaSna/Index.asp.

Burkhauser, R. V., Hérault, N., Jenkins, S., and Wilkins, Roger. What Has Been Happening to UK Income Inequality since the Mid-1990s? Answers from Reconciled and Combined Household Survey and Tax Return Data." February 2016. NBER Working Paper w21991.

Burkhauser, R. V., Feng, S., Jenkins, S. P., & Larrimore, J. 2011. Trends in United States Income Inequality Using the March Current Population Survey: The Importance of Controlling for Censoring. *Journal of Economic Inequality* 9(3), 393–415.

Burkhauser, R. V., Feng, S., Jenkins, S. P., & Larrimore, J. 2012. Recent Trends in Top Income Shares in the USA: Reconciling Estimates from March CPS and IRS Tax Return Data. *Review of Economics and Statistics* 94(2), 371–88.

Burkhauser, R. V., Larrimore, J., & Simon, K. 2012. A Second Opinion on the Economic Health of the Middle Class. *National Tax Journal* 61(1), 7–22.

Canberra Group (2011). *Handbook on Household Income Statistics*, second edition. Geneva: United Nations Economic Commission for Europe.

Cilke, J. 2014 The Case of the Missing Strangers: What We Know and Don't Know About Non-Filers. *Mimeo*

Congressional Budget Office. 2016. *The Distribution of Household Income and Federal Taxes, 2013.* Washington, DC: U.S. Government Printing Office.

Cunningham, C. R., & Engelhardt, G. V. 2008. Housing Capital-Gains Taxation and Homeowner Mobility: Evidence from the Taxpayer Relief Act of 1997. *Journal of Urban Economics* 63, 803–15.

d'Ercole, M. M., & Förster, M. 2012. The OECD approach to measuring income distribution and poverty: strengths, limits and statistical issues. In: D. J. Besharov and K. A. Couch (eds.), *European Measures of Income and Poverty: Lessons for the U.S.* New York: Oxford University Press, 27–58.

Dowd, T., McClelland, R., & Muthitacharoen, A. 2012. *New Evidence on the Tax Elasticity of Capital Gains: A Joint Working Paper of the Staff of the Joint Committee on Taxation and the Congressional Budget Office.* Joint Committee on Taxation JCX-56-12.

European Commission, International Monetary Fund, Organisation for Economic Co-operation and Development, United Nations, and World Bank. 2008. *System of National Accounts 2008.* Available online via: http://unstats.un.org/unsd/nationalaccount/docs/SNA2008.pdf.

Feldstein, M., & Yitzhaki, S. 1978. The Effects of the Capital Gains Tax on the Selling and Switching of Common Stock. *Journal of Public Economics* 9(1), 17–36.

Finkelstein, A., Hendren, N., & Luttmer, E. F. P. 2015. The Value of Medicaid: Interpreting Results from the Oregon Health Insurance Experiment. *NBER Working Paper 21308.*

Frick, J. R, Grabka, M. M., Smeeding, T. M., & Tsakloglou, P. 2010. Distributional Effects of Imputed Rents in Five European countries. *Journal of Housing Economics* 19(3), 167–79.

Gottschalk, P., & Danziger, S. 2005. Inequality of Wage Rates, Earnings and Family Income in the United States, 1975–2002. *Review of Income and Wealth* 51(2), 231–54.

Haig, R. M. 1921. The concept of income-economics and legal aspects, in R. M. Haig (ed.) *The Federal Income Tax*, Columbia University Press, New York: pp. 54–76.

Harris, B., & Moore, B.D. 2013. Residential Property Taxes in the United States. *Urban-Brookings Tax Policy Center Research Report*.

Kuznets, S. 1955. Economic Growth and Income Inequality. *American Economic Review*, 45(1): 1–28.

Larrimore, J., Mortenson, J. & Splinter, D. 2017. Household Incomes in Tax Data: Using Addresses to Move from Tax Unit to Household Income Distributions. *Board of Governors of the Federal Reserve System Finance and Economics Discussion Series 2017-002.*

Lundberg, J., & Waldenström, D. Forthcoming. Wealth Inequality in Sweden: What Can We Learn from Capitalized Income Tax Data. *Review of Income and Wealth.*

National Association of Realtors. 2016. *Median Sales Price of Existing Homes* [HOSMEDUSA052N], retrieved from FRED, Federal Reserve Bank of St. Louis https://research.stlouisfed.org/fred2/series/HOSMEDUSA052N/, October 19, 2016.

Piketty, T., & Saez, E. 2003. Income Inequality in the United States, 1913–1998. *Quarterly Journal of Economics* 118(1):1–39. Supplementary data updated to 2015 in June 2016. Available online via: http://elsa.berkley.edu/~saez/.

Piketty, T., Saez, E., & Zucman, G. 2016. Distributional National Accounts: Methods and Estimates for the United States. *NBER Working Paper 22945*.

Proctor, B.D., Semega, J., & M.K. Kollar. 2016. *Income and Poverty in the United States: 2015*. US Census Bureau Population Reports, P60-256. Washington, DC: U.S. Government Printing Office.

Roine, J., & Waldenström, D. 2012. On the Role of Capital Gains in Swedish Income Inequality. *Review of Income and Wealth* 58(3), 569–87.

Saez, E. 2016. Striking it Richer: The Evolution of Top Incomes in the United States (Updated with 2015 Preliminary Estimates). Available online via: http://eml.berkeley.edu/~saez/saez-UStopincomes-2015.pdf.

Saez, E., & Zucman, G. 2016. Wealth Inequality in the United States since 1913: Evidence from Capitalized Income Tax Data. *Quarterly Journal of Economics* 131(2), 519–78.

Shan, Hui. 2008. The Effect of Capital Gains Taxation on Home Sales: Evidence from the Taxpayer Relief Act of 1997. *Federal Reserve Board Working Paper 2008-53.*

Simons, H. C., 1938. *Personal Income Taxation: The Definition of Income as a Problem in Fiscal Policy*. Chicago University Press: Chicago.

Slemrod, J. 2016. Caveats to the Research Use of Tax return Administrative Data. *National Tax Journal* 69(4), 1003–1020.

Smeeding, T., & Thompson, J. 2011. Recent Trends in Income Inequality: Labor, Wealth and More Complete Measures of Income. *Research in Labor Economics* May, 1–49.

US Department of Treasury. 2016. *Taxes Paid on Long-Term Capital Gains, 1977-2013.* https://www.treasury.gov/resource-center/tax-policy/tax-analysis/Documents/Taxes-Paid-on-Long-Term-Capital-Gains.pdf

Veall, M.R. 2012. Top Income Shares in Canada: Recent Trends and Policy Implications. *Canadian Journal of Economics* 45(4), 1247–42.

Wolff, E., & Zacharias, A. 2009. Household Wealth and the Measurement of Economic Well-being in the United States. *Journal of Economic Inequality* 7, 83–115.

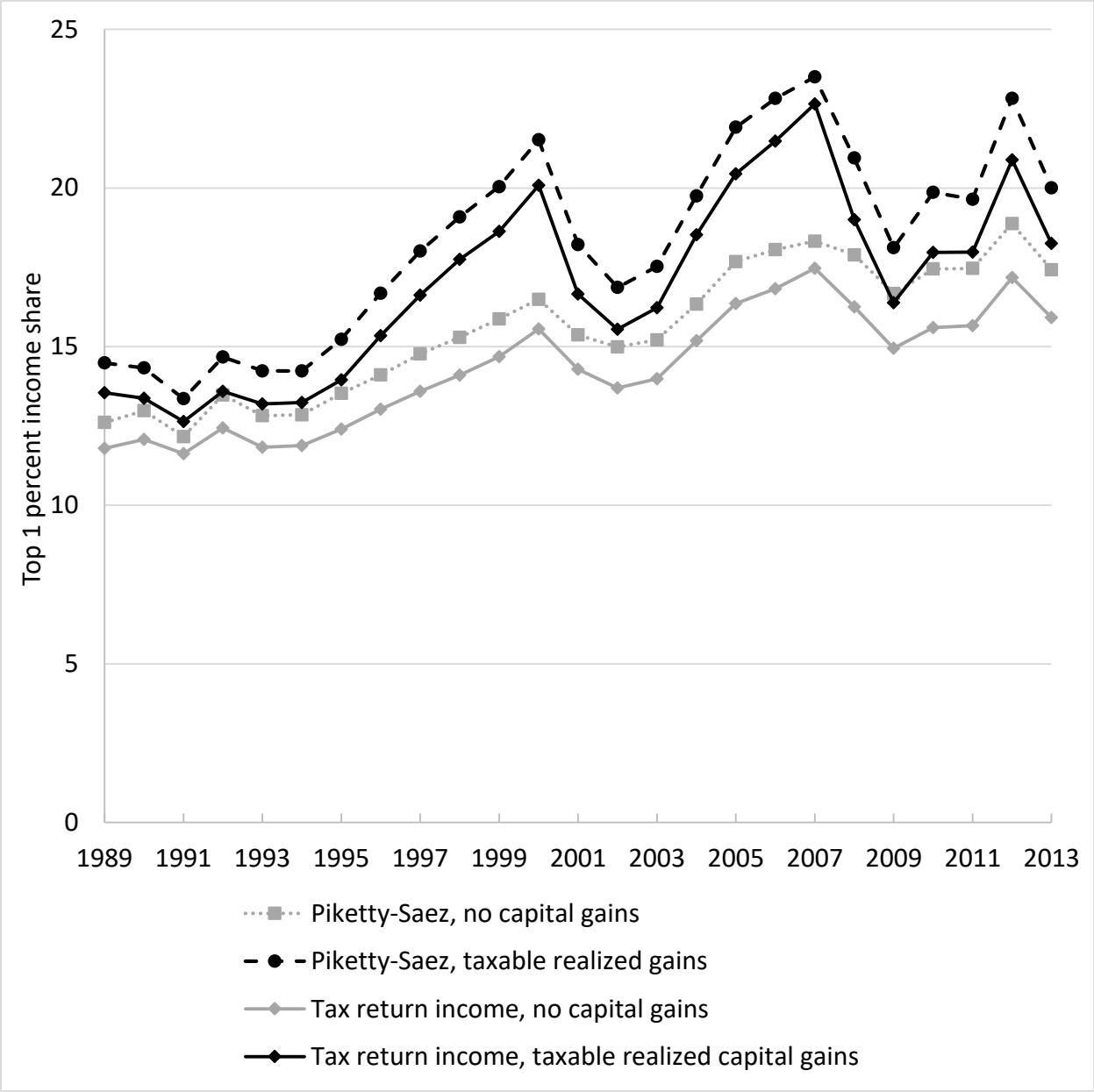
Wolff, E., Zacharias, A., & Masterson, T. 2012. Trends in living standards and inequality, 1959-2007. *The Review of Income and Wealth* 58(2), 475–99.

Table 1: Realized Capital Gains in 1997 through 2007, by length of time holding the asset before sale

Holding period	Percent of realized capital gains
less than 1 year	2.8
1 to 2 years	13.3
2 to 5 years	21.2
5 to 10 years	19.6
10 to 20 years	23.4
20 years or more	19.8

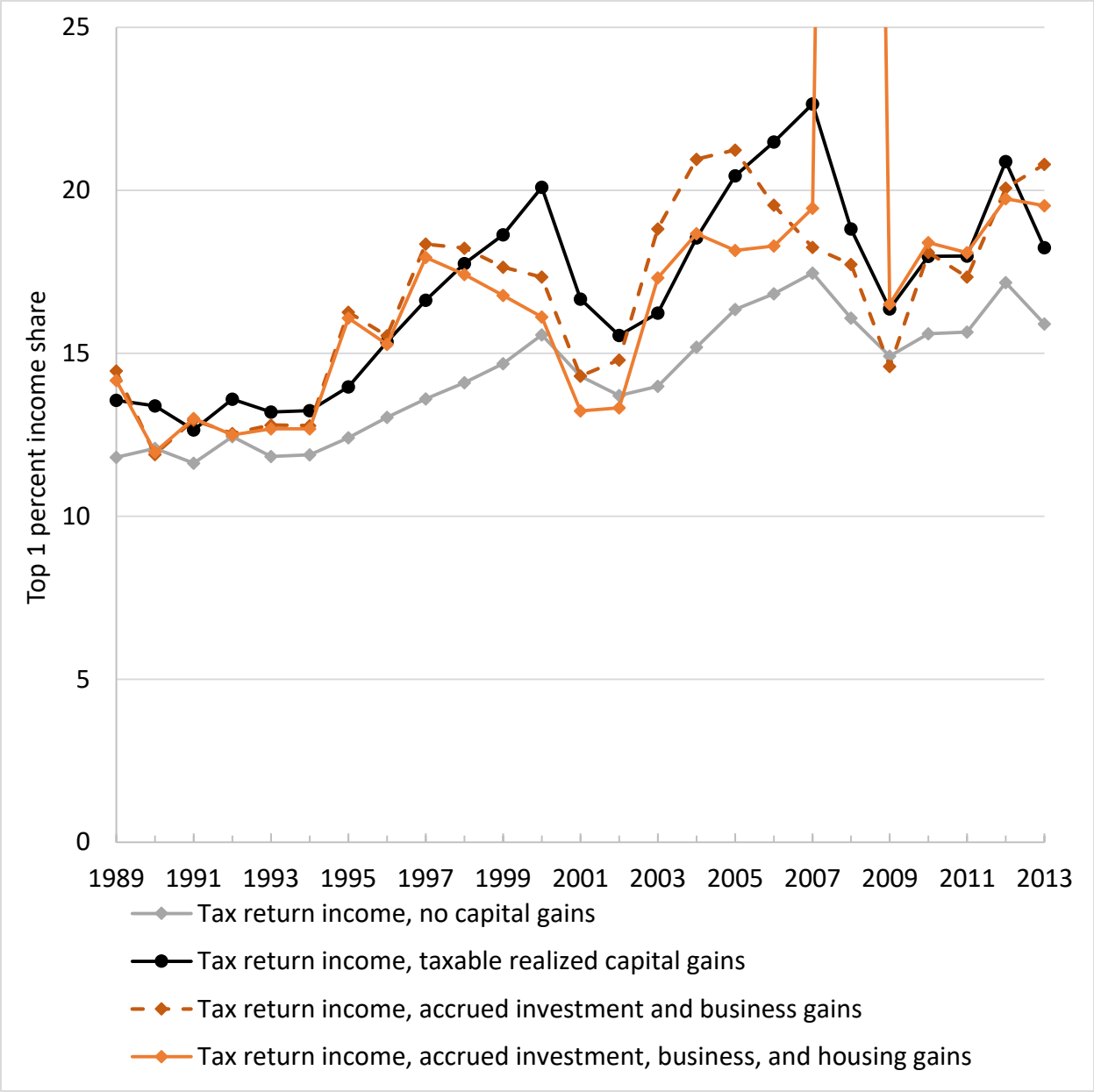
Source: IRS Sale of Capital Asset data, 1997 to 2007 (Schedule D)
Note: Short-term gains of unknown holding period are included as less than one year. Long term gains with unknown holding period are excluded.

Figure 1: Top 1% income share for market income and tax return income with and without realized taxable capital gains



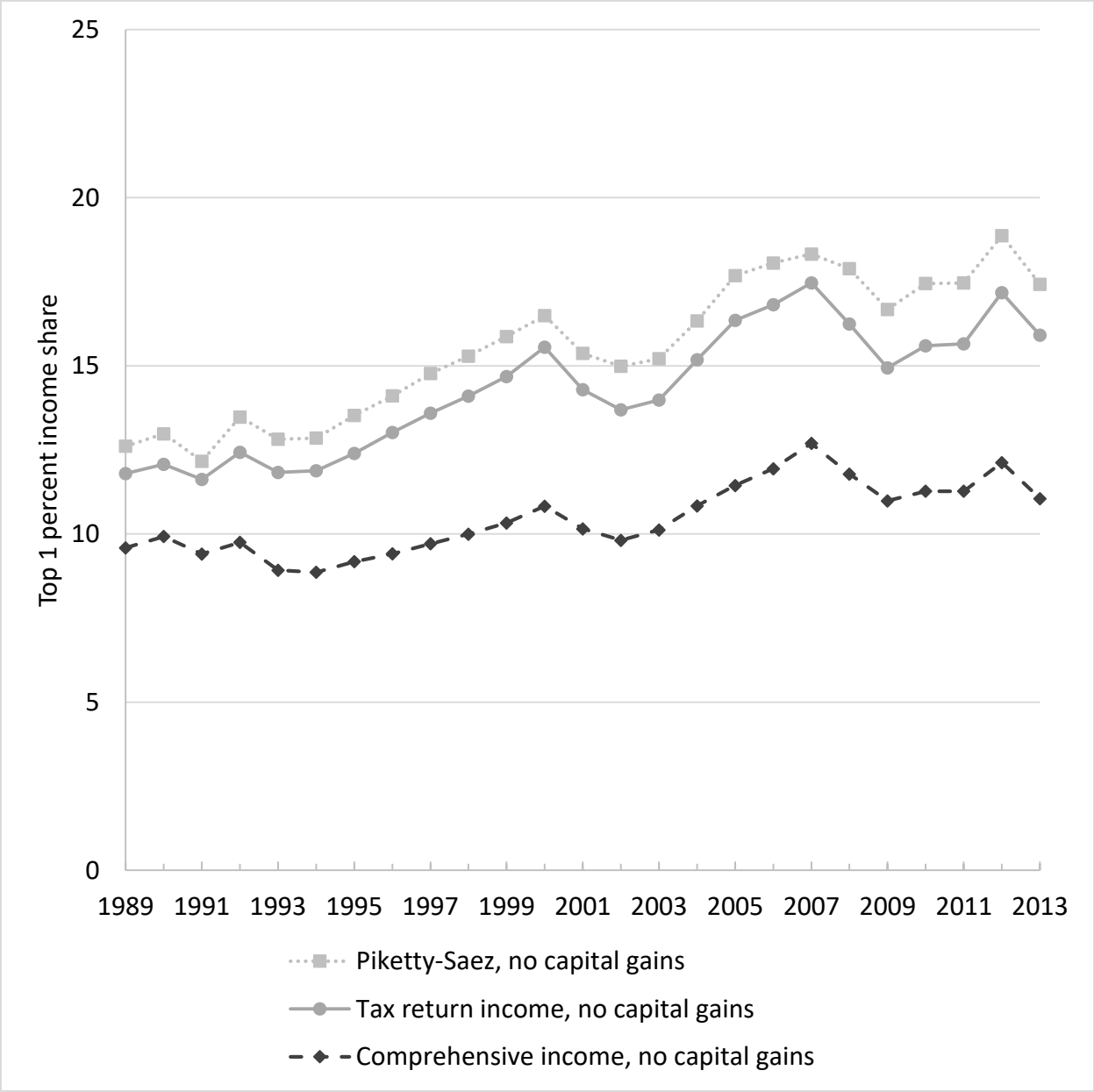
Source: Piketty and Saez (2003) and authors' calculations using IRS tax return data

Figure 2: Top 1%income share for tax return income with various treatments of capital gains



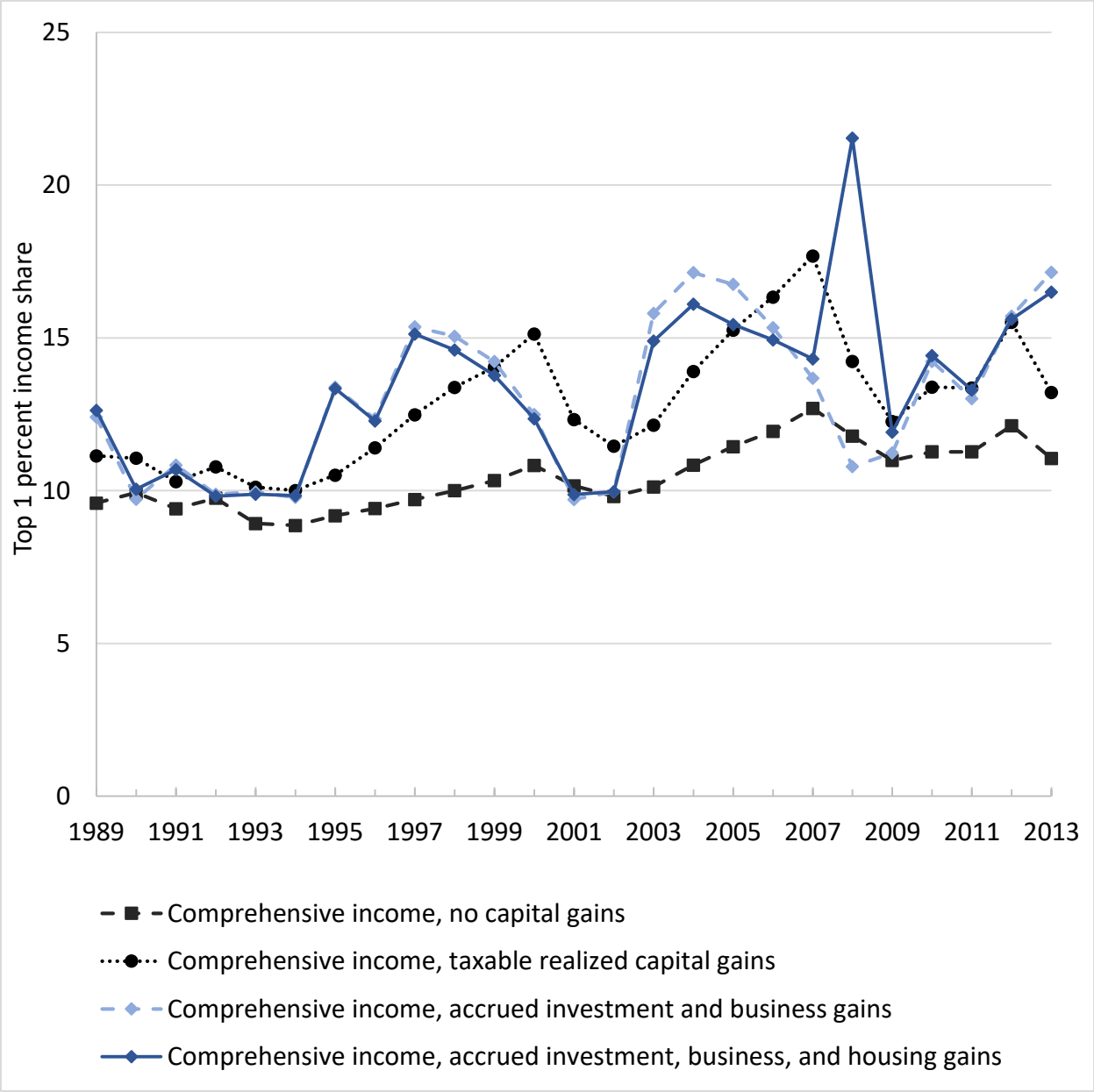
Source: Authors' calculations using IRS tax return data, the Survey of Consumer Finances, the American Community Survey, and the Decennial Census.

Figure 3: Top 1% income share for various income definitions, excluding capital gains



Source: Piketty and Saez (2003) and authors' calculations using IRS tax return data and the March CPS.

Figure 4: Top 1% income share for comprehensive income with various treatments of capital gains



Source: Authors' calculations using IRS tax return data, the Survey of Consumer Finances, the March Current Population Survey, the American Community Survey, and the Decennial Census.

Figure 5: Comparing top 1% income share using comprehensive income measures to that seen using market income with taxable realized capital gains

Source: Piketty and Saez (2003) and authors' calculations using IRS tax return data, the Survey of Consumer Finances, the March Current Population Survey, the American Community Survey, and the Decennial Census.

Data Appendix

1. Determining the population of tax units and the number of non-filers

The total number of tax units for this paper starts with the number of potential tax units in each year from Piketty and Saez (2003) and their subsequent updates. This count is based on Census Bureau data of the total number of individuals or married couples who are over the age of 20. The Census resident population of the United States includes people whose usual residence is in the 50 states and the District of Columbia and does not include members of the Armed Forces overseas, their dependents, or other U.S. citizens residing outside the United States. We then remove from the total count of potential tax units the dependent filers who are over age 20, since they do not represent independent entities. As a result, in each year we have somewhat fewer tax units than is observed by Piketty and Saez.

The total number of tax units that file a return in each year is the total number of non-dependent resident tax filers who submit a return covering the specific income year. This approach differs from that taken by Piketty and Saez in three ways. The first is that this approach removes tax filers who report their residence as being outside of the United States. The second is that it removes all dependent filers from the data, whereas Piketty and Saez include them. However, although we remove all dependent filers, removing those under age 20 does not impact the total number of (filing plus non-filing) tax units under their definition since these young dependent filers were not counted as potential tax units in their Census-based population count.[30]

The third way our count of the number of filing tax units differs from Piketty and Saez is that we focus on the year in which income is received rather than the year in which the tax return is filed. For example, if an individual files their tax return for income year 2000 in the spring of 2002, we treat this return as reflecting their 2000 income, whereas the data that Piketty and Saez use will consider this income with the other returns filed in 2002 that reflect 2001 incomes (the calendar year prior to filing). Additionally, if the same taxpayer files an amended return in the spring of 2003, then their income from 2000 would appear twice in the data used by Piketty and Saez. We correct this problem by reassigning late-filed returns to the income year for which the income was reported and use only the most recent tax return submitted.

The number of non-filing tax units in each year reflects the residual between the total number of non-dependent tax units for that year and the total number of non-dependent tax units who file a return.

2. Incorporating non-taxable income sources for comprehensive income

Most income of tax units is determined based on the incomes reported on annual tax return forms, such as the Form 1040. However, there are several cash and in-kind transfers and employee compensation that do not appear on tax returns but are captured by the Census Bureau's March Current Population Survey (CPS). The cash income items that are included in

[30] The problem with dependent filers under age 20 is particularly notable in 2007, when more tax returns were filed—including dependent filers—than there were potential tax filers under the Piketty and Saez (2003) approach.

the March CPS but not the IRS data are public assistance income (such as AFDC/TANF), supplemental security income, child support, education and financial assistance income, veteran's income, and worker's compensation. The in-kind resources, which we value at market value and include with income, are food stamps, school lunches, housing assistance, Medicare, Medicaid, and employer provided health insurance. (See Burkhauser, Larrimore, and Simon 2012 for a broader discussion of the decision to value health insurance resources at their market value.)

To incorporate these additional sources of income into the comprehensive income measure, we first construct tax units in the CPS data using the approach described in the main text, whereby all individuals over age 20 and their spouses are considered separate tax units (along with any dependent children that they may have who are under age 20). These tax units are then split into two groups based on whether any individual in the tax unit is over age 65. The split by age is intended to recognize that several of these items, such as Medicare, are substantially more prevalent among older adults. Within the two age groups, the tax units are then arrayed based on their income (incorporating only those components that appear on tax returns) and split into 100 centiles based on their position in the income distribution within their age group. A similar procedure is undertaken on the IRS side to create centiles of the income distribution within the two age groups in the tax data.

Using the March CPS, we compute the mean value of each income source that does not appear in the tax data for each centile of the population within the two age groups. This mean value is then assigned to individuals in the same centile and age group in the IRS data and added to the resources of that tax unit when computing comprehensive income.

3. Imputed rents for owner occupied housing

The Haig-Simons income measure includes imputed rents along with the other income sources described above. To incorporate imputed rents, we start with the aggregate level of imputed rents in each year from the Bureau of Economic Analysis' Gross Domestic Product calculations. We then distribute these imputed rents to homeowners in proportion to their estimated home values (calculated using the procedure from section 4.A of this appendix and as discussed in the main text) to reflect that homeowners with more expensive homes will also have higher levels of imputed rents.

In most years, the inclusion of imputed rents reduces the top 1% income share by approximately one-half percentage point. This reflects that imputed rents (and housing wealth) is less concentrated among the top 1% than is the rest of comprehensive income. The inclusion of imputed rents has a larger impact on the top 1% income share in 2008. In this year, the fuller income measure substantially reduces the spike in top 1%income shares, since the addition of this income source has a larger relative impact on the overall levels of income than it does in other years. With the inclusion of imputed rents as part of income, the top 1% income share still spikes in 2008 but not to the same degree as seen in the series that exclude these values from income.

4. Determining asset values for use in accrued capital gains calculations

A. Housing Values, itemizers

Among taxpayers who itemize their taxes, estimates of home values start with line 6 of Schedule A of the IRS Form 1040 – the deduction for real estate taxes paid. Recognizing that some individuals may not itemize every year, for individuals who remain in the same zip code for three consecutive years and itemize in years *t-1* and *t+1*, but not in year *t*, their real estate tax payments in year *t* are assigned the average of their real estate tax payments in the two adjoining years. The real estate tax is interpolated for individuals who fail to itemize for two years but remain in the same zip code all four years. After these adjustments, taxpayers who remain in the same zip code are assumed to have the same amount of real estate tax for up to three additional years. Though this procedure may slightly understate the tax paid, it provides a close approximation and helps identify additional home-owners.

To convert real estate tax payments to home values, we use self-reported property tax payments and home values from the Decennial Census (1990 and 2000) and from the American Community Survey (ACS) (2005 through 2013). In these surveys, respondents report the estimated market value of their house as well as the property tax payments.[31] Using these data, we compute the ratio of the average property value to the average property tax payment for each county using the following approach:

(1) For years since 2005, when ACS data is available, the ratio of property values to property tax payments is taken from the analysis by Harris and Moore (2013) of county-level property taxes using this data. Not all counties have data available using the one-year ACS, so a hierarchical approach is taken, where one-year ACS data is used for counties where it is available. If the one-year ACS data is not available for a county then the three-year data is used, and if the three-year ACS data is not available then then five-year data is used. When multi-year ACS data is used for a county, the value-to-tax ratio for a year is the multi-year estimate with a midpoint closest to the year of interest.

(2) For the two decennial census years (1990 and 2000), the ratio of property values to property tax payments is computed using the IPUMS public use microdata sample. Since IPUMS data does not report the county, we compute the property tax and property values at the Public Use Microdata Area (PUMA) level and then calculate the county-level result as the population-weighted average of overlapping PUMAs using the PUMA-County crosswalk files from the Missouri Census Data Center.[32]

(3) For years between decennial census years, the ratios in each county are computed as the weighted average of the two decennial censuses—so the ratio for a county in 1991 will equal 90 percent of the ratio from that county in 1990 plus 10 percent of the ratio

[31] For the Decennial Census, property tax payments are reported in ranges, and we use the mid-point of each range and 1.5 times the topcode value ($4,500 in 1990 and $9,100 in 2000) for the highest category. We similarly assign topcoded property values (above $600,000 in 1990 and above $1 million in 2000) 1.5 times the topcode threshold.

[32] See, http://mcdc.missouri.edu/allabout/geo_pumas.shtml.

from that county in 2000. Similarly, for years between the 2000 Decennial Census and the start of the ACS data in 2005, the ratio for each county is computed as the weighted average between the 2000 Decennial Census and the 2005 ACS.

The county-level ratios of home values to property tax payments are then assigned to each zip code (since zip codes are reported on tax forms whereas counties are not) using the county-zip crosswalk file from the Department of Housing and Urban Development.[33] In the case where a zip code includes multiple counties, the zip code is assigned the results from the county in which the largest fraction of the zip code's residents reside.

Individual-level property values are then computed as:

$$Value_{i,t} = Tax_{i,t} * \frac{\overline{Value}_{county,t}}{\overline{Tax}_{county,t}}$$

where $Value_{i,t}$ is the taxpayer-level home value for itemizers in year *t*, $Tax_{i,t}$ is the taxpayer-level property tax payment in year *t*, $\overline{Value}_{county,t}$ is the county-level average home value in year *t*, and $\overline{Tax}_{county,t}$ is the county-level average property tax payment in year *t*.

This approach for valuing houses using the individual-level property tax payments and the county-level home value to property tax ratio uses an assumption that even if different counties assess properties in different ways, within a county there is general uniformity in the relationship between property taxes and home value. There are two notable ways in which this assumption could fail. The first is if counties offer tiered property tax rates, such as exempting a fixed dollar amount of property values from taxation. This would result in low-priced homes paying lower property taxes than would be seen if all homes were taxed at the mean rate for the county. As a result, our approach will underestimate the value of low-cost homes in these areas and overestimate the value of high-cost homes in these areas. The second way in which this assumption could fail is when counties fail to have horizontal equity in their property tax rates. The most notable example of this occurs in California, where Proposition 13 limits the magnitude of reassessments except when a sale or construction occurs. As a result, a long-tenured homeowner will typically have a lower tax bill than one who purchased their home more recently – even if the market value of the two homes is identical. Our approach for valuing homes using property taxes in this case will result in an overestimate of the value of homes of new homeowners (whose property taxes are above the mean for a home with their value) and will underestimate the value of homes for long-term homeowners.

Each of these limitations will not affect the total estimated home value in a county, but will impact who is estimated to have higher value homes. To the extent that it results in an underestimate or overestimate of home values of taxpayers in the top 1% of the distribution, our method could therefore result in an underestimate or overestimate of our top income shares. However, despite this limitation, we believe that this approach—and the potential error that it introduces—is superior to the alternative approach for estimating home values based purely on

[33] See, https://www.huduser.gov/portal/datasets/usps_crosswalk.html.

one's position in the income distribution, which can similarly result in an underestimate or overestimate of top income shares. Estimated home values are capped at $100 million. Although advertised prices for homes for sale sometimes exceeds this amount, in most case such large values likely reflect multiple homes which may be in areas with different tax rates. In addition, such large homes may be unique or idiosyncratic homes that reflect the tastes of the current owner. The values of such homes may not change in the same way as average homes in the area.

B. Housing Values, non-itemizers

Whereas nearly all taxpayers in the top centile of the income distribution who own a home itemize their taxes, many lower-income taxpayers do not. As a result, property tax payments are not available for these non-itemizers to compute their estimated home value.

Since the non-itemizers are assumed to be outside of the top centile of the income distribution, it is not necessary to know precisely who the non-itemizing homeowners are in order to calculate top income shares, only the total amount of their real estate assets. In each year, we calculate the total home values that are assigned to itemizing taxpayers from section 3.A above and compare it to the total value of real estate owned by households and non-profits serving households from the Integrated Macroeconomic Accounts. The residual home values that are not assigned to itemizers are then assigned to non-itemizers. The assignment of this home value is done quasi-randomly, but ensuring that the estimated home values do not push any non-itemizers into the top 1% of the distribution.

C. Equities and business assets

The value of equities and business assets are computed using data from the Survey of Consumer Finances (SCF) using a statistical match based on their receipt of dividends, business ownership, and income. First, since the SCF aggregates individuals to the Primary Economic Unit (PEU) rather than the tax unit, it is necessary to map PEUs into tax units—which we do by assuming that the PEUs that contain multiple tax units are evenly distributed through the distribution.

Tax units in both the SCF and IRS data are then categorized into four groups based on whether they have 1) at least $100 of dividend income, 2) at least $500 of rental and/or small business income or losses, 3) both dividend and rental/small business income above these thresholds, or 4) neither dividend nor rental/small business income above these thresholds. These categories of separation were chosen to reflect that tax units with dividends and/or small business income are more likely to have capital asset holdings than those who do not. Within these four categories, tax units are then arrayed based on their total tax return income, excluding capital gains, and assigned a centile based on their position in the income distribution within their category group.

For each centile of the income distributions within the four groups outlined above, we compute the average value of all equity assets (including stocks and mutual funds in both taxable and non-taxable accounts) and the average value of all privately held businesses based on the reported values from the SCF. Since the SCF is only conducted every three years, for years between the SCF years, the asset values for each centile are set as the weighted average of the two nearest SCF years. Tax units in the IRS data are then imputed to have equity and business asset values

from this SCF data based on their dividend/business income characteristics and based on their centile of the income distribution of those with these characteristics.

5. Determining rates of return for use in accrued capital gains calculations

A. Housing rate of return

Among taxpayers who own a home and itemize their deductions, the rate of return on the real estate assets are based on year-over-year change in the CoreLogic Home Price Index for their county of residence (based on the zip-county crosswalk described in section 3.A above). For counties where CoreLogic produces a county-level home price index, the yearly change in this county-level value represents the rate of return for housing assets. For counties where a county-level home price index is not available, the state-level value is used. County-level rates of return are available and used for approximately two-thirds of all zip codes.

Among non-itemizing homeowners, since we are unable to determine precisely which non-itemizers are homeowners and the home values are assigned to the non-itemizers collectively, the rate of return on housing in each year is determined using the national-level change in the CoreLogic Home Price Index.

B. Equities and private business assets rate of return

The rate of return for equities and private business assets are both assumed to be uniform for all individuals in each year. This approach is consistent with the assumption made by Saez and Zucman (2016) when estimating asset values using information from tax return data. For all equity holdings, the rate of return in each year is estimated to be the year-over-year change in value of the S&P 500 index. For private business assets, the rate of return in each year is estimated to be the implied rate of return on non-corporate business holdings of households and non-profits serving households from the Integrated Macroeconomic Accounts. This implied rate of return is computed as the revaluation of equity in non-corporate business in the calendar year divided by the prior year end-of-year value of non-corporate business holdings. The annual rates of return on publicly traded investments and non-corporate business investments are available in Appendix Table A1.

6. Computing accrued capital gains

Accrued capital gains for each asset type are determined based on:

$$CG_{i,t} = AssetValue_{i,t} * ROR_{i,t}$$

Where $CG_{i,t}$ is the capital gains from an asset type for the taxpayer, $AssetValue_{i,t}$ is the total value of the type of asset by the taxpayer from section 3 above, and $ROR_{i,t}$ is the rate of return that the asset receives from section 4. The total capital gains and losses observed through this approach from equities, privately held business assets, and housing can be seen in Appendix Table A2.

Appendix Table A1: Annual rate of return on selected assets and percentage change in realized capital gains, 1989-2013

	Annual change in realized gains in AGI	S&P 500	Non-corporate business (IMA)	CoreLogic House Price Index
1989	-5.3	27.3	5.4	8.6
1990	-19.6	-6.6	1.2	2.8
1991	-9.8	26.3	-2.2	-1.2
1992	13.5	4.5	-1.4	0.0
1993	20.2	7.1	3.2	1.5
1994	0.3	-1.5	5.4	2.2
1995	17.9	34.1	4.7	1.9
1996	44.7	20.3	3.5	2.9
1997	39.9	31.0	11.0	3.3
1998	24.8	26.7	6.9	7.0
1999	21.4	19.5	7.1	7.3
2000	16.6	-10.1	10.8	9.8
2001	-45.8	-13.0	4.1	9.5
2002	-23.1	-23.4	7.6	8.8
2003	20.4	26.4	9.5	9.8
2004	54.4	9.0	18.8	14.7
2005	38.3	3.0	16.3	16.7
2006	15.7	13.6	5.1	6.6
2007	15.8	3.5	1.2	-4.9
2008	-46.1	-38.5	-13.4	-13.2
2009	-47.1	23.5	-15.1	-11.6
2010	49.6	12.8	7.3	-0.6
2011	2.6	0.0	6.1	-3.8
2012	60.0	13.4	9.1	3.9
2013	-21.1	29.6	11.8	11.1

Notes and Sources: The IMA returns to non-corporate business is from the Federal Reserve Integrated Macro Accounts. The House Price Index is from the CoreLogic data. Realized capital gains are positive capital gains in Adjusted Gross Income as reported in U.S. Treasury (2016).

Appendix Table A2: Aggregate components of comprehensive income excluding capital gains, by year (millions of dollars)

	Tax return income	In-kind benefits and in-kind transfers	Imputed rents	Cash transfers not on tax forms	Net federal income and payroll taxes	Total comprehensive income, excluding capital gains
1989	3,430,883	255,491	388,900	83,621	-612,939	3,545,956
1990	3,643,423	286,747	417,600	94,240	-637,819	3,804,191
1991	3,806,921	334,818	441,400	113,560	-646,301	4,050,398
1992	3,948,283	357,164	468,000	115,538	-681,323	4,207,662
1993	4,053,800	399,620	494,500	120,446	-717,112	4,351,255
1994	4,242,183	462,369	530,300	123,781	-756,903	4,601,730
1995	4,501,782	460,754	566,200	124,046	-816,400	4,836,383
1996	4,776,646	483,192	599,900	128,320	-897,004	5,091,054
1997	5,113,017	472,672	636,400	124,517	-983,471	5,363,134
1998	5,461,559	478,730	681,400	131,725	-1,054,782	5,698,632
1999	5,818,032	466,158	726,300	133,039	-1,161,347	5,982,182
2000	6,252,010	499,125	777,700	136,373	-1,287,993	6,377,216
2001	6,368,447	581,242	838,200	155,858	-1,193,537	6,750,209
2002	6,375,469	623,170	878,500	147,114	-1,105,856	6,918,398
2003	6,544,863	711,293	922,900	160,261	-1,065,843	7,273,475
2004	6,980,430	822,567	980,300	169,055	-1,163,490	7,788,862
2005	7,430,291	888,182	1,057,900	179,450	-1,276,025	8,279,798
2006	7,949,609	934,543	1,123,100	167,348	-1,390,209	8,784,390
2007	8,313,861	815,181	1,156,100	174,187	-1,499,614	8,959,715
2008	8,565,351	925,889	1,199,300	187,358	-1,326,720	9,551,177
2009	8,200,689	952,798	1,223,700	193,966	-1,196,391	9,374,762
2010	8,581,932	1,087,330	1,228,700	212,092	-1,290,200	9,819,854
2011	8,891,446	1,150,446	1,249,200	220,647	-1,304,001	10,207,738
2012	9,460,494	1,208,549	1,278,800	237,488	-1,457,248	10,728,082
2013	9,564,954	1,210,827	1,316,500	244,343	-1,605,376	10,731,247

Appendix Table A3: Aggregate taxable realized capital gains and accrued capital gains, by year (millions of dollars)

	Taxable realized capital gains	Accrued gains from equities	Accrued gains from private business holdings	Accrued gains from housing	Total accrued capital gains
1989	143,375	374,817	158,720	558,103	1,091,640
1990	112,318	-102,226	34,848	192,230	124,852
1991	99,990	459,260	-64,910	-125,863	268,487
1992	116,684	86,942	-41,992	43,837	88,787
1993	141,959	164,528	98,325	102,654	365,507
1994	140,510	-36,547	178,482	185,366	327,301
1995	168,519	1,085,854	157,076	150,466	1,393,396
1996	247,524	868,991	135,003	267,678	1,271,672
1997	349,244	1,688,884	474,103	342,906	2,505,892
1998	439,097	1,754,015	321,355	724,461	2,799,830
1999	533,461	1,481,983	379,167	858,777	2,719,926
2000	625,335	-873,777	646,774	1,309,376	1,082,373
2001	308,381	-1,260,669	268,978	1,343,613	351,922
2002	237,392	-2,281,133	540,071	1,351,524	-389,538
2003	294,469	2,584,675	723,634	1,749,096	5,057,405
2004	480,882	878,429	1,525,546	2,937,706	5,341,681
2005	659,782	316,078	1,541,127	3,796,336	5,653,542
2006	783,769	1,547,306	550,784	1,746,824	3,844,914
2007	913,321	432,130	150,922	-757,618	-174,566
2008	432,703	-4,469,765	-1,543,132	-2,139,661	-8,152,558
2009	221,068	2,575,485	-1,615,769	-1,874,638	-914,921
2010	358,822	1,323,223	721,629	-225,141	1,819,711
2011	376,560	-360	627,954	-651,420	-23,826
2012	617,324	1,641,379	962,233	445,349	3,048,961
2013	475,326	3,916,562	1,301,911	2,128,191	7,346,664