TAX FOUNDATION
SPECIAL REPORT
Aug. 2014
No. 224

Income Data Is a Poor Measure of Inequality

By Alan Cole
Economist

Key Findings

- IRS income data is collected in order to raise revenue as directed by Congress, which means it is not necessarily well-suited for other purposes, like measuring equality in our society.
- The average taxpayer's income changes dramatically throughout his lifetime; the average tax return for an 18- to 25-year-old shows about $15,000 in adjusted gross income where an average tax return for someone between ages 55 and 64 shows above $80,000.
- College students, particularly, comprise a very large number of low-income taxpayers.
- Incomes go considerably farther in some places than in others. Much of the narrative about rural states being poorer is mistaken.
- Much capital income—especially capital income in tax-free middle-class retirement accounts—goes uncounted in income data, heavily distorting the measurement and making people appear poorer than they are.
- Thomas Piketty's income inequality data leaves out $19 trillion of pension assets, which are yet to be attributed to any individual.

Introduction

The Internal Revenue Service (IRS) collects data on the incomes of individual taxpayers, because the amount of tax owed is based on income. The IRS releases some of this data, in part for social science research. Frequently, however, the data is used to show more than it actually should. It's reasonable to wonder how income is distributed among taxpayers, but social scientists are cavalier about the limitations of income data, especially income as defined for tax purposes.

Studies of income distribution tend to get their data from two sources: the IRS Statistics of Income or the U.S. Census American Community Survey. For example, the Census Bureau offers data on household income by quintile in one of its tables, like in Table 1.

Table 1. Household Income Quintiles, 2012

Quintile	Income Range
Lowest	$0-20,599
Second	$20,600-39,764
Third	$39,765-64,582
Fourth	$64,583-104,096
Highest	$104,097+

Source: U.S. Census Income Table H-1.1

The IRS offers similar data broken down by tax unit instead of by household. Both of these data sets are useful, and both have considerable strengths. The IRS data comes from a large sample, and for obvious reasons, it is very robust when it comes to all kinds of taxable income. The Census data comes from a smaller sample size and isn't as detailed as IRS data on income, but it is considerably more detailed with regard to other characteristics of the households surveyed.

While both of these data sets have their uses, they are not good for quantitatively measuring the inequality in standard of living among Americans. This weakness stems not from any fault of the workers at the IRS or the Census Bureau but rather from the nature of the data itself. Income over one year is simply a very poor proxy for standard of living.

This is in part because of the lack of context—no person is defined by a single year's worth of income data—and in part due to the weaknesses of income measurement itself. Income is not all there is to class and mobility in America—not by a long shot. The faux precision of quintiles and Gini coefficients abstracts away context and data issues, leaving researchers with the impression that they understand a great deal more about people's lives than they actually do.

One flaw in income data is that it tells you about only one year in someone's life, while people's lives play out over much longer time periods. Another is that differences in regional price levels—especially rent—make nominal income data a poor measure of people's standard of living. A final flaw in income data is that it is collected for tax purposes and not for the purpose of social science. Many types of income are not counted at all, or counted in ways that don't reflect economic realities.

The purpose of this paper is not to address whether income inequality has increased or decreased. Rather, it is to show how the aforementioned flaws can affect the data substantially, produce counterintuitive results, and ultimately have adverse effects on policies based on income.

Income Varies Dramatically Over Life Cycles

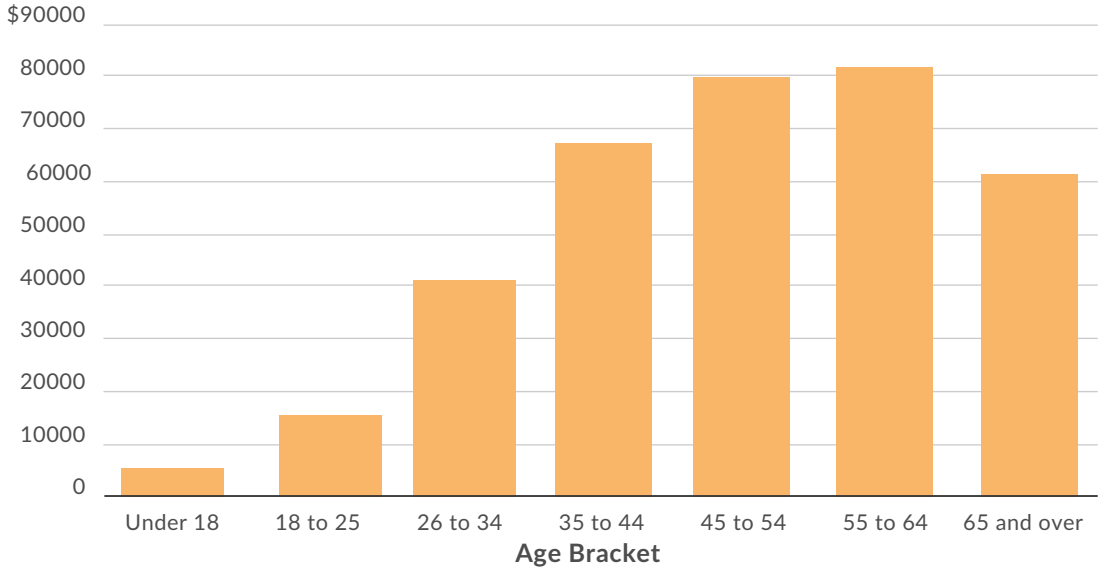
Income data is almost always reported in annual terms. This makes a great deal of sense to the IRS. The IRS is tasked with collecting revenue on an annual basis, and it collects revenue based largely on what people earned in the past year.

This turns out to be of limited usefulness in describing people's general wellbeing. People can shift spending money between years by saving, drawing down savings, or borrowing. They also plan their careers on horizons of decades or more. One year's data tells us very little about someone's life.

The general arc of income data—often known as the earnings-age profile—is an inverted-U shape with respect to age.

Figure 1. Income Changes Over the Course of an Individual's Life

Adjusted Gross Income by Age Bracket



Source: IRS, *Statistics of Income Division, (July 2013)*

Grouped by age (Figure 1, above), this data looks reasonable. Income is low below the age of 26 when many Americans are still in school. Then it rises with age as they accumulate savings and work experience. Finally, it comes back down as they start to retire. An American earning the average adjusted gross income (AGI) for his age ends up in all five of the AGI quintiles throughout her lifetime.[1]

1 According to the IRS Statistics of Income Tax Stats for 2011, the 80th percentile AGI was approximately $79,838. The average tax filer between ages 55 and 64 earned $81,859, putting him in the top quintile.

Income data over only a year misses this trend—effectively, you end up comparing people to older or younger versions of themselves. There is a substantial mathematical inequality between a 21-year-old with a $16,000 income and a 56-year-old with an $80,000 income. Among those two, five-sixths of the income accrues to the 56-year-old. Yet it would be a mistake to draw a larger narrative about haves and have-nots from these two average individuals. Any model of social inequality that can be driven by perfectly average individuals is unrealistic. Average people can't be drivers of any meaningful inequality, virtually by definition.

Longitudinal studies that follow the same taxpayer over several years show a very different story from a one-year snapshot. A large portion of those taxpayers who earn low incomes in a particular year will move on to earn higher incomes as they grow older. In 2010, the Tax Foundation's Robert Carroll studied IRS data on a set of taxpayers' returns from 1999 to 2007. Of those taxpayers that were in the lowest quintile in 1999, a majority—57.5 percent of them—were in a higher quintile by 2007.[2] In other words, while some low-income taxpayers remained low income, the majority of them did not.

The picture we get when we compare people by age—or look at taxpayers over time—is one that fits with personal experience. Americans gradually build their careers, acquire skills, and figure out how to best participate in the economy. This creates high measured income inequality over one-year periods but also high mobility over the long term.

For example, if you observed the tax data for surgeons, you would find very high degrees of inequality. Through their twenties, surgeons are in medical school and earning very little. When they finally become residents, they earn moderate amounts. In their prime, with a proven track record and solid experience, they can earn $400,000 or more. The income inequality among surgeons of different ages is staggering.

Incomes Are Lowest in College Towns

Young people in general, and students specifically, make income data almost unusable for some purposes. For example, one might want to find data on the poorest places in America. It may seem that the best way to do this is through examining household income. But that gets you some odd results that don't really reflect "poverty" as it's typically imagined. This is because contrary to most people's expectations, America's lowest incomes are actually found in college towns.

2 Robert Carroll, *Income Mobility and the Persistence of Millionaires, 1999 to 2007*, Tax Foundation Special Report No. 180 (June 21, 2010), http://taxfoundation.org/article/income-mobility-and-persistence-millionaires-1999-2007.

Table 2. The Fifteen Lowest-Income Places in the United States

Census Place	Number of Households	Median Household Income (2012 Dollars)	Mean Household Income (2012 Dollars)	Higher Education
Boone, North Carolina	6221	$16,447	$33,173	Appalachian State University
Carbondale, Illinois	9601	17743	36495	Southern Illinois University
Athens, Ohio	6497	18428	41163	Ohio University
East St. Louis, Illinois	10612	19278	27732	
Statesboro, Georgia	9850	20751	32586	Georgia Southern University
Opa-locka, Florida	5163	20757	30512	
East Cleveland, Ohio	8074	20797	29280	
Opelousas, Louisiana	6110	20983	33058	
University CDP (Hillsborough County), Florida	16657	20992	28893	University of South Florida
Selma, Alabama	7812	22076	40092	
Milledgeville, Georgia	5839	23580	39824	Central Georgia Technical College, Georgia College, Georgia Military College
Prichard, Alabama	8714	23726	32582	
State College, Pennsylvania	12178	24104	45138	Pennsylvania State University
Pullman, Washington	10151	24125	47808	Washington State University
Oxford, Ohio	5730	24211	49458	Miami University

Source: U.S. Census Bureau, American Community Survey (Selected Economic Characteristics, 5-year Estimates).
Note: minimum 5,000 households.

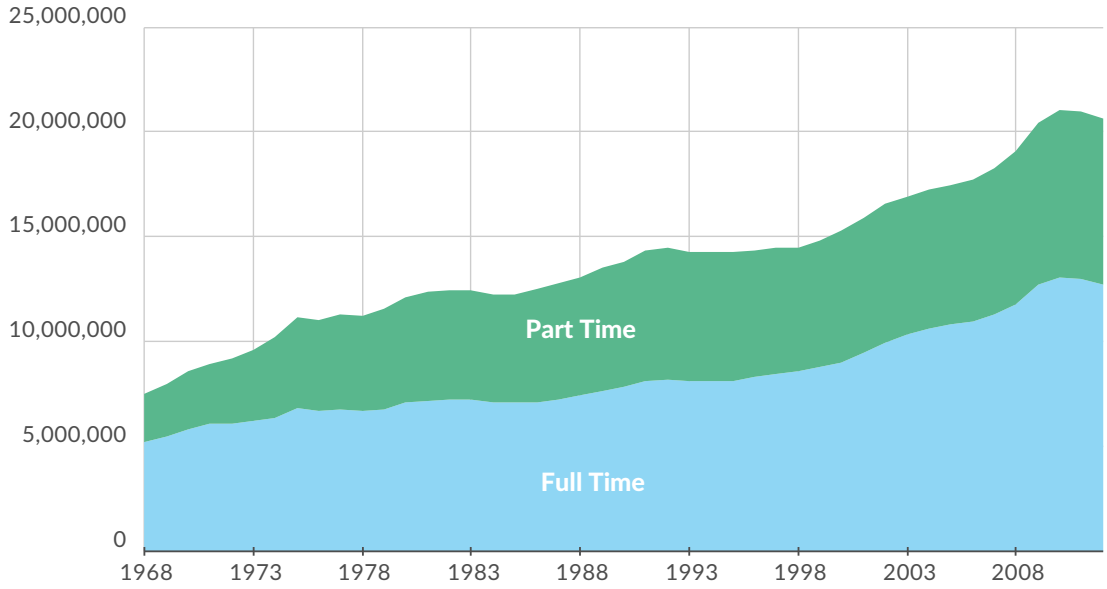
This table understates the magnitude of the effect. It takes results from the American Community Survey, whose economic supplement excludes students in dormitories. In other words, students skew economic data so strongly that off-campus students alone are a dominant determinant of household income. The Census Bureau has itself studied the issue and found that 51.8 percent of all off-campus college students not living with relatives have been counted as living below the poverty level.[3] While college juniors living in apartments might have some budget constraints, it is highly implausible to think the majority of them are appropriate targets for the War on Poverty.

By no means are these small data curiosities that can be brushed off. The number of students in America is large and growing fast with over 20 million students enrolled in post-secondary education in the United States.

3 U.S. Census Bureau, Alemayehu Bishaw, *Examining the Effect of Off-Campus College Students on Poverty Rates*, SEHSD 2013-17 (May 1, 2013), http://www.census.gov/hhes/www/poverty/publications/papers-bishaw.html.

Figure 2. Post-Secondary Enrollment Has Increased Over Time

Post-Secondary Enrollment by Year



Source: National Center for Education Statistics, Digest of Education Statistics, Table 303.10.

As the number of students has grown (Figure 2, above), the effect on income data has become quite substantial. About 140 million individual tax returns are filed in the typical year,[4] and many of them come from the population of students. Most of these students appear to the IRS—for now—as low-income taxpayers.

While this growth in higher education skews income data considerably, it is nonetheless a good thing for young Americans. The returns to education are substantial; the U.S. Census Bureau found that in 2012, households headed by someone with a bachelor's degree (but no graduate degree) had a median income of $80,549. The median household headed by someone with a doctorate ($116,983) or a professional degree ($129,588) was even better off. [5]

Those households headed by people with a bachelor's would be in the fourth quintile. Those headed by people with doctorates or professional degrees would be in the highest quintile. But many individuals holding those degrees spent time in the lowest quintile in order to earn them.

Economists, researchers, and journalists often consider low incomes a sign of a lack of opportunity. At times that is undeniably true. But there is yet a lot of opportunity in America, and, paradoxically, that opportunity can be greatest in places where incomes are lowest.

4 Internal Revenue Service, Brett Collins, *Projections of Federal Tax Return Filings: Calendar Years 2011-2018*, IRS Statistics of Income Bulletin (Winter 2012), http://www.irs.gov/pub/irs-soi/12rswinbulreturnfilings.pdf.

5 U.S. Census Bureau, Current Population Survey, Annual Social and Economic Supplement, Table HINC-01, *Selected Characteristics of Households by Total Money Income in 2012*, https://www.census.gov/hhes/www/cpstables/032013/hhinc/hinc01_000.htm.

America's Substantial Disparities in Cost of Living

Economists often talk about nominal and real data. When they compare dollar-denominated quantities over time, they often adjust them for inflation. If *nominal* wages rise over time, but the prices of goods rise over time by the same amount, then people haven't really gotten richer at all; real wages have remained constant.

A similar practice can be followed with comparing dollar-denominated quantities across space; nominal price levels are different in different places. Adjusting for these differences is called price level parity. In April, the U.S. Bureau of Economic Analysis released regional price parities (RPP) for the first time, allowing for study of price level differences among locations within the United States.

The highest RPPs in the United States will surprise no one; they center around the Bay Area on the West Coast and New York City on the East Coast. The San Francisco-Oakland-Fremont Metropolitan Statistical Area, for example, has a regional price parity of 123.5, meaning that for the basket of goods the BEA measured, the San Francisco area is about 23.5 percent more expensive than the rest of the nation.[6]

This solves some puzzles that would otherwise be difficult to solve. The conventional wisdom about Oakland is that it has some substantial economic challenges to overcome—probably more than most places in America. That conventional wisdom is not borne out by its median household income of $51,683, which is close to the national average of $53,046.[7] However, deflate Oakland's income by its regional price parity, and you get a RPP-adjusted income of only $41,849, considerably below the national average. After applying even some basic adjustments to nominal incomes, the conventional wisdom about Oakland begins to make sense again.

Price parity adjustments help solve a number of puzzles at the state level. In a series on interstate migration, Tax Foundation economist Lyman Stone found that people relocate, on net, not to the places with higher nominal incomes but to the places with higher *price-adjusted* incomes.[8] In other words, high rent is legitimately unpleasant and people consider that a factor when choosing where to live.

This also helps solve another state puzzle—the observation that rural, low-income states frequently vote against redistributive economic policies, seemingly against their own interests. In 2004, political analyst Thomas Frank wrote a bestselling book, *What's the Matter with Kansas?*, attempting to explain this phenomenon in his home state, which persistently had nominal incomes below the national average. The book focused on cultural issues that shape the state's political climate.

6 Bureau of Economic Analysis, Bettina Aten, Eric Figueroa, & Troy Martin, *Notes on Estimating Regional Price Parities by 16 Expenditure Categories: 2005-2009*, http://bea.gov/papers/pdf/notes_on_estimating_the_multi_year_rpps_and_appendix_tables.pdf.

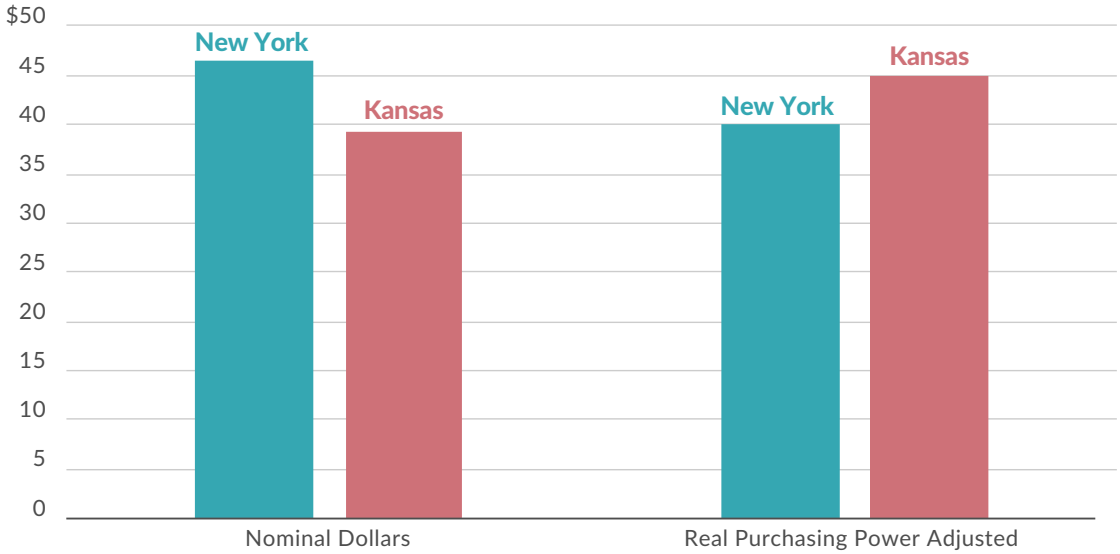
7 U.S. Census Bureau, American Community Survey (Selected Economic Characteristics, 5-year Estimates).

8 Lyman Stone, *New State-Level Price Data Shows Smaller State Real Income Differences*, Tax Foundation Tax Policy Blog, May 23, 2014, http://taxfoundation.org/blog/new-state-level-price-data-shows-smaller-state-real-income-differences.

An RPP-adjusted comparison tells a different, though not mutually exclusive, story. While Kansas has had persistently lower incomes than, for example, New York, it also has a much lower cost of living. In the BEA's working paper on regional price parities, Kansas comes out ahead in RPP-adjusted income (see Figure 3).

Figure 3. Kansas Has a Higher Average Income than New York after Adjusting for Purchasing Power

Real Purchasing Power of Per Capita Personal Income in New York and Kansas, 2009 (Thousands of Dollars)



Source: Bureau of Economic Analysis, Aten, Figueroa, & Martin, *Notes on Estimating the Multi-Year Regional Price Parities by 16 Expenditure Categories: 2005-2009.*

This is not to say that Kansas is necessarily a better place to be than New York. Both states have economic challenges and economic strengths. Nor is it to say that people should move from New York to Kansas for a superior standard of living. Some people thrive best in the borough of Manhattan on the Hudson River. Other people thrive best in the city of Manhattan on the Kansas River. These are not judgments that can be determined by a government bureau or a social scientist. Rather, these are judgments made by individuals with knowledge of their own personal circumstances.

So what's the matter with Kansas? Why does it vote as if it's not a poor state? Probably many reasons—but one of these is that on the whole, it's not particularly poor in the first place. A cursory look at Leawood or Lenexa could attest to that. A cursory look at nominal income data could not.

Inconsistent or Absent Measurement of Non-Wage Income

The largest problem of all with income data is that it isn't even a good measure of income. There's a simple reason for this. The IRS is the only government agency that rigorously requires you to report your income. But some of your income is not taxable, and some of it is not even reported to the IRS.

The problem starts with capital gains, which are measured only when realized. This creates extreme spikes in measured capital income, when in truth the capital income was accrued over many years. If you invest in stock at age 25 and then cash it out at age 65 to fund your retirement, all forty years of capital gains will be counted at age 65.

This same inconsistent measurement process occurs with shares of S corporations as well. If you are a small business owner, the growth of your business's equity value is not recorded as income until you sell it.

This definition of capital gains income works well for the IRS's purposes. The IRS can't realistically spend its time assessing the value of every asset you own every year. From the IRS's perspective, it's much better to tax you only when you realize gains on your asset.

However, this definition of capital gains income gives us a very confusing—and overly unequal—perception of what people's capital income actually is like. This skewed distribution shows up strongly in longitudinal studies, just like the life cycle effects discussed previously. Robert Carroll's study of income mobility looked at a sample of nine years and found that 50 percent of the millionaires—people with a million dollars of taxable income for at least one year—were millionaires *only* for one year. For many of the millionaires, this was an artifact of capital gains measurement; the volatility of millionaire status dropped substantially if capital gains data were excluded. Carroll concluded, "Millionaires are a highly transient group of taxpayers, and it appears that the realization of capital gains is at least one explanation."[9]

When you look at only one year of income data, you lose these nuances. Some people look like they earned a million dollars in a year, when in fact it may have taken them decades to accrue those gains. At the same time, other people look like they're living a more modest lifestyle, when in fact they have substantial unrealized capital income.

In other words, just like education creates massive distortions at the low end of the income distribution, capital gains realizations create massive distortions at the upper end of the income distribution.

But the middle class is not without its own distortions, and those distortions aren't minor.

Middle-class Americans have quite a great deal of capital income that the IRS simply never sees. Capital income on owner-occupied homes is largely exempt—both in terms of imputed rent earned on the home and in terms of capital gains below a minimum threshold.[10] Owner-occupied housing represents a capital stock of $20.2 trillion,[11] which provides Americans with a steady stream of housing as well as potential capital gains.

9 Carroll, *supra* note 3.
10 Currently this threshold is $250,000, or twice that for married couples.
11 Federal Reserve Board, *Financial Accounts of the United States, Flow of Funds, Balance Sheets, and Integrated Macroeconomic Accounts*, Tables B.100 & R.100 (June 5, 2014), http://www.federalreserve.gov/releases/z1/Current/z1.pdf.

Also exempt are middle-class retirement savings vehicles. U.S. households have $19.8 trillion worth of pension entitlements.[12] Every single dollar in a 401(k), Traditional IRA, or employer-sponsored plan, public or private, has *never once* been counted by the IRS as *anyone's* income. In 2013, for example, pension funds had assets of $18.9 trillion,[13] all of which was earned by somebody. None of that money has yet showed up on an individual tax return.

In 2006, Cato Institute Fellow Alan Reynolds criticized income inequality data compiled by the French economist Thomas Piketty and the American economist Emmanuel Saez along these grounds: "In recent years, an increasingly huge share of the investment income of middle-income savers is accruing inside 401(k), IRA and 529 college-savings plans and is therefore invisible in tax return data."[14] Piketty and Saez responded blithely, as if unaware of the scale of the problem: "Even the small point on 401(k) s is conceptually mistaken: pension income is reported on tax returns when withdrawn during retirement and hence returns on pension funds are implicitly included in our income measure."[15]

Piketty and Saez's response is only true for a minority of Americans—those seniors who have actually reached retirement age. It is not true for the majority of Americans, nor for the tens of trillions of dollars they have invested for retirement. Piketty's dismissal of nearly $20 trillion as a "small point" about his income measure is absurd. A measure that "implicitly" counts the money only on a decades-long lag is not a good measure at all.[16]

Adjusting for these issues with capital income paints an entirely different picture. Last year, Philip Armour, Richard Burkhauser, and Jeff Larrimore published a paper in the *American Economic Review* that imputed the *accrued* capital income, as opposed to *realized* capital income, among quintiles. Using a consistent definition of accrued capital income, the authors found that such a measure dramatically reduced income inequality overall. Furthermore, with this measure, income growth among the quintiles has been equal since 1989.[17]

Changes in income inequality depend a great deal on what one considers income. Measures based on IRS data—which exclude tax-free retirement accounts—will invariably create a distorted picture where the very tax breaks that enrich the middle class, like 401(k) accounts, paradoxically make that middle class look far poorer.

12 Federal Reserve Board, *Financial Accounts of the United States, Flow of Funds, Balance Sheets, and Integrated Macroeconomic Accounts*, Tables B.100 & R.100 (June 5, 2014), http://www.federalreserve.gov/releases/z1/Current/z1.pdf.

13 Towers Watson, *Global Pension Assets Study 2014* (Feb. 5 2014), http://www.towerswatson.com/en-US/Insights/IC-Types/Survey-Research-Results/2014/02/Global-Pensions-Asset-Study-2014.

14 Alan Reynolds, *The Top 1%...of What?*, Wall Street Journal, Dec. 14, 2006, http://online.wsj.com/news/articles/SB116607104815649971.

15 Thomas Piketty & Emmanuel Saez, *Response by Thomas Piketty and Emmanuel Saez to: The Top 1% ... of What?*, *http://eml.berkeley.edu/~saez/answer-WSJreynolds.pdf*.

16 Nonetheless, Piketty and Saez's "implicit" counting of retirement saving is still better than using Census data. The American Community Survey only asks about regular sources of income, not one-time withdrawals from retirement. Therefore, it permanently excludes 401(k)s and IRAs.

17 Philip Armour, Richard V. Burkhauser, & Jeff Larrimore, *Deconstructing Income and Income Inequality Measures: A Crosswalk from Market Income to Comprehensive Income*, 103 American Economic Review 173-177 (2013).

Conclusion

IRS income data is collected for the purpose of raising revenue annually in the manner that Congress directs. It was not intended to be a measure of one's overall wellbeing. In the absence of better data, some social scientists are tempted to use IRS data that way.

This is a mistake. Income data has massive confounding factors; not minor technical nitpicks, but big glaring issues so plain and so relevant that they can be expressed in terms of the lives of ordinary people. People develop professionally with age. People go to college. People think about where rents are high and where they are low. People save in retirement accounts.

Income data would be a reliable measure of social inequality if it weren't *distorted by virtually every major decision people make in their lives*. Income data, out of context, leads us to conclusions so absurd that the entire project of dividing people up into quintiles may be an intellectual dead end. The dollar-denominated sum of certain classes of market transactions is not enough to identify suffering or plenty.

The United States has a progressive tax code. The primary intellectual basis for the progressivity is the idea that people with lower incomes are more in need of money than people with higher incomes. Overall, this is undeniably true. But it is far less true than often imagined. And it is particularly untrue among everyday Americans whose IRS-defined income is a poor proxy for their social wellbeing.

Because of the deadweight losses in taxes, errors in redistribution matter a great deal. Marginal tax rates discourage work, saving, and investment. If money is collected to provide struggling people with help to get by, that's one thing. But all too often, the limits of income data result in socially-nonsensical redistribution. A reasonable person would not say Oakland, California is substantially better off than Green Bay, Wisconsin. And yet, Oakland shoulders a far higher per-household tax burden.[18] A reasonable person would not say a construction worker is clearly better off than a business school student. And yet, it is the latter who benefits from progressive income taxes and refundable tax credits at the expense of the former. Use data unreasonably and you will get unreasonable results.

As an instrument of the federal government, the IRS is best when used for its intended purpose: collecting revenue. It is considerably less effective at creating social justice, which is not something easily determined using a Form 1040 alone. Efforts to fight social inequality would be best undertaken by humane institutions with well-defined purposes and local knowledge of the problems they are designed to handle—not a large centralized bureau built to extract revenue on a mass scale.

18 Gerald Prante & Andrew Chamberlain, *Estimating Federal Tax Burdens for Major City Areas, Counties, and U.S. Congressional Districts*, Tax Foundation Working Paper No. 2 (Mar. 22, 2007) (calculating that Oakland's total average tax burden for the calendar year 2004 was $24,818. Green Bay's was $17,226).

The Tax Foundation is a 501(c)(3) non-partisan, non-profit research institution founded in 1937 to educate the public on tax policy. Based in Washington, D.C., our economic and policy analysis is guided by the principles of sound tax policy: simplicity, neutrality, transparency, and stability.

Editor, Donnie Johnson
Designer, Dan Carvajal

Tax Foundation
National Press Building
529 14th Street, NW,
Suite 420
Washington, DC
20045-1000

202.464.6200

taxfoundation.org