

Advertisement

A black banner with orange bokeh lights. On the left is the Tableau logo. In the center, the text reads 'TOP TEN Big Data TRENDS FOR 2017'. On the right, a white box contains the text 'GET THE WHITEPAPER'.

Become a member [Renew my subscription](#) • [Sign up for newsletters](#)



MacArthur Foundation

How Two Economists Got Direct Access to IRS Tax Records

By [Jeffrey Mervis](#) | May. 22, 2014, 2:00 PM

Raj Chetty of Harvard University and Emmanuel Saez of University of California (UC), Berkeley, created a **big media splash** last summer with a study showing that social mobility—the income status of adult children relative to their parents—correlates with where the children grew up. The study, based on an analysis of millions of U.S. tax records that had been largely off-limits to researchers, has fed the public perception that the American dream of equal opportunity for all may be fading. It also bolstered the reputations of the two young superstars—each has received the top prize for economists under 40 and a MacArthur “genius” award. And it has left their colleagues wondering how they pulled off such a feat.

“It was very entrepreneurial of them to get access to the data, which is not normally available,” says Gary Solon, an economist at Michigan State University in East Lansing who has done pioneering work on social mobility using small data sets from surveys, the traditional approach to studying the topic. “You need the energy and perseverance and connections. My guess is that it was probably some combination of skill and luck.”

Solon’s hunch is right, according to the U.S. IRS, which described the unusual arrangement in a series of e-mail exchanges with *ScienceInsider*. Solon and his fellow researchers aren’t driven simply by idle curiosity: **Access to government administrative data** that can be linked to surveys may hold the key to unlocking the causal factors behind social mobility, an important but poorly understood phenomenon. But a host of issues, from privacy to cost, are complicating efforts to tap such troves.

SIGN UP FOR OUR DAILY NEWSLETTER

Get more great content like this delivered right to you!

A tweak in the tax code

The story behind the Chetty/Saez paper begins in 1987, when the IRS started requiring taxpayers to list the Social Security number of every dependent listed on tax returns. The new rule, a trivial component of a major overhaul of the U.S. tax system, was designed to stop parents from claiming imaginary dependents that would lower their tax bills. And it worked: Several million fewer dependents were listed on 1987 tax returns.

Chetty was 9 years old at the time and living in India, and Saez was a teenager in France. After a meteoric rise through academia—Chetty was granted tenure at UC Berkeley at the tender age of 27, for example—they earned their academic spurs with a series of theoretical and empirical papers on how various government policies influence human behavior.

Tax cheats hadn't been on their radar. But a few years ago they realized the change in the IRS tax forms made it possible, for the first time, to link millions of children with their parents' tax records. Those children could then be followed into adulthood and, thus, be part of a study on intergenerational mobility. After persuading the Treasury Department's Office of Tax Policy that such a study could shed light on whether local tax and spending policies affect social mobility, the team was given the chance to work directly with tax records.

The researchers began by examining tax returns filed in 1996. They identified a core sample of nearly 10 million children born between 1980 and 1982 (14- to 16-year-olds). The researchers then tracked the children until approximately age 30 and compared their family incomes with those of their parents. They calculated the parent's income by averaging total family income over 5 years, from 1996 to 2000; for the adult children, they measured income in 2011 to 12. The team then ranked the incomes of both parents and adult children in relation to their peers and divided each group into quintiles. In a last step designed to tease out geographical differences in mobility, they assigned the children to the city in which they were living at age 16; in all, they study included 731 localities spanning the entire country.

To simplify their results, the researchers calculated the chances that a child from a family whose income was in the lowest quintile in the 1990s will have jumped into the top quintile by age 30. The national average was 7.5%, but the percentage varied greatly by geography. In San Jose, California, 12.9% of the children who grew up there managed to make the big jump (putting it at the top of a mobility list of the 50 largest U.S. cities). In contrast, those from Charlotte, North Carolina, brought up the rear, at 4.4%.

A second study, drawn from the same database, found mobility has remained relatively constant over time in recent decades. Specifically, they reported that the probability a child born in the bottom fifth would leap to the top fifth was 8.4% for children born in 1971, compared with 9.0% for those born in 1986. The conclusion, based on a younger cohort born in the 1990s, draws upon both the children's income and a proxy involving any amount of college attendance. Chetty says that the two estimates "exhibit very similar trends" and that the researchers therefore rely on income "as our main estimate." But many social mobility researchers question the value of such a proxy and say the results about trends are much less persuasive than the work on regional variations.

Getting their hands dirty

The process that allowed Chetty and Saez to work directly with the IRS records was routine, the two researchers say. "We got access to the tax data through a standard call for research proposals from IRS," Saez told *Science* in an e-mail. Declining to answer other questions, Saez acknowledged that the team, which included Harvard's Nathaniel Hendren and UC Berkeley's Patrick Kline, took a path rarely trodden by researchers. "They unfortunately have very little funding," he wrote about the IRS, "and hence can only accommodate a relatively small number of researchers."

A research solicitation that IRS issued in the fall of 2011 attracted 51 proposals, according to Barry Johnson, head of the special studies branch of the Statistics of Income division. Some 19 were accepted, Johnson says, and 16 studies were actually carried out. And the vast majority of researchers supported by the IRS were required to follow a protocol that allowed them to use the information without actually handling the microdata itself.

"We were given a dummy data set, with random numbers, to test our program," explains David Grusky, a sociologist at Stanford University in California and director of its Center on Poverty and Inequality. "Once we're confident the program is working, we ship it off to the IRS and someone there does the run. After checking to make sure no confidential information is included, they send the output back to us. And we shuffle back and forth until the project is done. It's a little cumbersome, but it works."

Such an arrangement is far from optimal, however, say scientists not involved in the IRS research program. "It's a bit awkward, a bit clunky, to get dummy data to debug your programs," explains Miles Corak, an economist at the University of Ottawa who has helped develop data sets on social mobility for the Canadian government. "The problem is that you don't get to dance and play with the data, and someone else runs it."

needed data,” Johnson says. Once that decision was made, the academics agreed to “submit to fingerprinting and a complete background check, undergo training in the proper protection of administrative data, and be subject to the same rules and penalties that apply to any IRS employee.” They also worked under the supervision of the Treasury Department; one employee, Nicholas Turner, was even listed as a co-author on one of the key papers.

At the same time, IRS required the authors to receive prior approval of any papers or presentations based on their analysis of the restricted data. “The IRS does not in any way attempt to influence findings,” Johnson writes. “The review is limited to ensuring that the data have been described and used correctly. [That is] a standard feature of peer review.”

That policy does not exist at the government’s de facto statistical agency, the U.S. Census Bureau. Ron Jarmin, who manages the Census Bureau’s research and methodology programs, says “we do not impose editorial control of the research product.” The agency does make sure that any personal data have been “deidentified” before they are made available to researchers, he adds. But beyond that, he says, the agency “doesn’t have the time and resources” to do such a vetting, nor does it see any need to do so. “As a statistical agency, we value their output,” he says about the collaborations with researchers.

Can it be replicated?

The IRS says it has a long-standing interest in the scholarly analysis of social mobility as a way to assess tax policy. But some social scientists say that the agency didn’t really recognize the value of outside collaborations until Alan Krueger, a noted Princeton University economist, became chief economist at the Treasury Department in 2009.

Krueger, who also served as chair of the White House Council of Economic Advisers before returning to Princeton in August 2013, says he supported academic use of the agency’s data. But he doesn’t take credit for the 2011 solicitation. And he notes that tight budgets limit the number of such collaborations. The IRS could do more, he adds, if somebody else footed the bill.

“My own view is that the IRS should charge researchers to cover the cost of accessing data, the same way that the Census Bureau does,” Krueger says. (His reference is to the fees charged to researchers who use the agency’s research data centers, a network of 14 secure sites around the country.)

Unless and until that happens, however, social scientists will have to be content with applauding Chetty and Saez and dreaming about what they might do if they could get ahold of such data. “For the purposes of measuring intergenerational mobility in the United States, it’s an amazing data set,” Solon says.

***Correction, 23 May, 12:41 p.m.:** *This Insider has been revised to clarify where the federal employee who oversaw the Chetty-Saez research project is employed.*

***Correction, 27 May: 3:47 p.m.:** *A reference to a cross section of children used as a sample in the first study has been removed, and information about the two estimates for income used in the second study has been added.*

See also:

- [The science of inequality](#)
- [The IGE: Anatomy of a Mobility Score](#)
- [What Surveys Have Told Us About U.S. Social Mobility](#)
- [It’s Already on File: How Administrative Records Can Help Assess Mobility](#)



Jeffrey Mervis

Jeff tries to explain how government works to readers of *Science*.

[✉ Email Jeffrey](#)

Please note that, in an effort to combat spam, comments with hyperlinks will not be published.

[♥ Recommend](#) [🔗 Share](#)

Sort by Best ▾



Join the discussion...

[@defensetax.com](#) • 3 years ago

But many social mobility researchers question the value of such a proxy and say the results about trends are much less persuasive than the work on regional variations. The IRS does not in any way attempt to influence findings. The agency does make sure that any personal data have been "deidentified" before they are made available to researchers

Thanks

IRS Tax Attorney @ <http://defensetax.com/>

1 ^ | ▾ • Reply • Share ›

[Unka_George](#) • 3 years ago

At last! An attempt to light a match and stop cursing the darkness. Not only does the IRS data illuminate the correlation between the economic status of the same gender parent and the child, it also offers the opportunity to quantify the value of a college education and existence/magnitude of the STEM shortage. It is critical this analysis is performed before any more shots in the dark are fired (as we are about out of ammunition). We must have data driven and not archaic ideologically based socioeconomic policies if the Republic is to endure and prosper.

1 ^ | ▾ • Reply • Share ›

[efalken](#) • 3 years ago

They have a lot of resources, and so should anticipate sophisticated analysis (they should understand what that means). Allowing selected individuals from institutions that can give them favors suggests a bias.

^ | ▾ • Reply • Share ›

[✉ Subscribe](#) [D Add Disqus to your site](#) [Add Disqus](#) [Add](#) [🔒 Privacy](#)

Got a tip?

[How to contact the news team](#)

Advertisement



Advertisement



Related Jobs

Flowminder Foundation/WorldPop Program is hiring Global Experts (Principal Investigators) in the UK

Southampton, Hampshire | £70,000-100,000 per annum

The Flowminder Foundation/WorldPop Program is hiring Global Experts (Principal Investigators), a Sampling & Survey Design Expert, and Data Analysts.

Employer: Flowminder

Faculty Positions available in the School of Environmental Science and Engineering

Shenzhen, Guangdong (CN) | Highly competitive salaries and benefit packages will be provided.

Applications are invited for faculty positions at all ranks in the School of Environmental Science and Engineering at SUSTech.

Employer: Southern University of Science and Technology (SUSTech)

Vice President for Academic Programs and Dean

San Francisco, CA | \$150,000-\$200,000 per annum


Latest News

1. Five-day fasting diet could fight disease, slow aging
2. These tiny camera lenses can 'see' like an eagle
3. How the battle lines over CRISPR were drawn
4. Want to repel mosquitoes? Don't use citronella candles
5. Another Alzheimer's drug flops in pivotal clinical trial


Most Read

1. Updated: U.S. House passes controversial bill on NSF research
2. Study retraction reignites concern over China's possible use of prisoner organs
3. U.S. Senate bill aims to make sure federal scientists aren't 'muzzled'
4. Cheap plastic film cools whatever it touches up to 10°C
5. The marches for science, on one global interactive map


Sifter




Brains of demented, former soccer pros show classic signs of CTE
Feb. 16, 2017




More than 700 species are already feeling the strain from climate change
Feb. 14, 2017



Trump freeze blocks bee from joining endangered species list
Feb. 10, 2017



Hearing loss is common in the United States—even for the young
Feb. 9, 2017



From insect bites to broken ankles: See your state's most unusually 'common' injury
Feb. 9, 2017

More Sifter

