Washington Center
for Equitable Growth

Where do the beneficiaries of the Affordable Care Act live?

A Technical Appendix

Carter C. Price   April 2015

Data

The primary data sets used in this analysis are the Statistics of Income, or SOI, from the U.S. Internal Revenue Service, the Current Population Survey March Supplement, and the U.S. Census Bureau's population estimates.

The SOI data contain various statistics about income tax filings across the United States. The data contain counts and averages by income group for areas as geographically fine as zip codes though the groups become coarser for smaller areas. For the purposes of this work, I will use county level data. Because this information is only on tax filings, some people—most notably those below the tax filing threshold (about $10,000 for a single person)—are excluded from this data.

For the Current Population Survey or CPS, I use the Center for Economic and Policy Research extracts. The CPS is nationally representative and has individual level details about people including demographics, geographic characteristics, and income information. The CPS has detailed characteristics about individuals, families, and households.

The Census Bureau produces population estimates by county that can be used to provide accurate estimates for the total population of an area. The census estimates are intended to count everyone and so can be used to correct the population counts in the SOI which will not include non-filers. I assume the difference in population between the Census estimates and the IRS's SOI to be non-filers and that the income for this segment of the population is below the filing threshold.

Methods

The first step to adjusting the CPS is to get a comparable unit of analysis. I constructed synthetic tax units in the CPS using relationship information as described in Piketty and Saez as well as in Burkhauser. Tax units can be single, married filing jointly, or, very

rarely, married filing separately. Because the CPS provides no way to identify married couples that would file separately, I assume that all couples file jointly.

Single people over the age of 20 are flagged as single filers and married people of any age are flagged as married filers. People under the age of 20 are assumed to be dependents and are assigned to their parents when the information is present.

Next, I determine the taxable income by summing the CPS's income information on wages and salary, self-employed farm work, self-employed non-farm work, dividends, interest, and rentals. This excludes some types of income but does manage to capture most taxable income for the vast majority of tax units. Notably, Capital gains are not included in this income information but this does not impact the matching because the SOI contains information both with and without capital gains.

For each income group and county in the SOI, I use the following equations to develop an income distribution that is continuous almost everywhere (the resulting probability density function could be discontinuous at each income group's boundary where the groups).

I assume that there exists a function

$f(q) = ae^{cq}$ such that $1 = \int_{q_0}^{q_1} f(q)dq$ and $\bar{q} = \int_{q_0}^{q_1} qf(q)dq$

I solve these equations to get:

$$1 = \int_{q_0}^{q_1} ae^{cq}dq = \frac{a}{c}e^{cq_1} - \frac{a}{c}e^{cq_0} \rightarrow$$

$$a = \frac{c}{e^{cq_1} - e^{cq_0}}$$

And

$$\bar{q} = \int_{q_0}^{q_1} aqe^{cq}dq = \frac{a}{c^2}\left(e^{cq_1} \cdot (cq_1 - 1) - e^{cq_0} \cdot (cq_0 - 1)\right) \rightarrow$$

$$c = \frac{a}{c\bar{q}}(e^{cq_1} \cdot (cq_1 - 1) - e^{cq_0} \cdot (cq_0 - 1)).$$

I used a fixed point method to approximate a and c for each income group by county. This produces an exponential distribution that reproduces the income information from the SOI for each county and income group. These distributions are then used to adjust the income groups of tax units. For $X \in [0,1]$:

$$X = \int_{q_0}^{q} ae^{cq}dq = \frac{a}{c}e^{cq} - \frac{a}{c}e^{cq_0} \rightarrow$$

$$q(X) = \frac{\ln\left(X \cdot \frac{c}{a} + \frac{a}{c}e^{cq_0}\right)}{c}$$

This distribution results in a function to provide estimates for the income of tax units. To convert tax unit income estimates into estimates of the poverty level, we need to match each income with an estimated family size. The SOI has limited data on household size and composition and so the CPS will be used to augment the estimates for the number of people in a tax unit. Specifically, I looked at the number of people within each tax unit by income using the CPS. Table 1 has the resulting distribution for people per tax unit by income level.

TABLE 1

Distribution of Tax Unit Size by Income

	$1 under $10,000	$10,000 under $25,000	$25,000 under $50,000	$50,000 under $75,000	$75,000 under $100,000	$100,000 under $200,000	$200,000 or more
1	54.82	43.44	35.84	27.82	15.94	9.36	4.79
2	27.15	37.49	26.68	26.31	28.31	28.45	29.03
3	8.16	7.88	14.33	15.72	17.91	19.30	19.47
4	5.19	5.89	12.31	16.09	21.66	25.67	29.44
5	3.05	3.28	6.60	8.68	10.06	11.64	12.61
6	1.54	1.81	4.02	4.93	5.67	5.28	4.40
7	0.44	0.74	1.23	1.90	1.93	1.47	1.30
8	0.16	0.39	0.33	0.75	0.95	0.43	0.34
9	0.09	0.25	0.04	0.43	0.30	0.21	0.11
10	0.00	0.13	0.05	0.18	0.15	0.03	0.05

Source: Author's calculations based on data from Current Population Survey

For each random sample, I draw two random numbers between zero and one that are used to sample from the income and population distributions. To estimate the share of the population eligible for Medicaid, I assess the number of people in tax units that fall below 138 percent of the Federal Poverty Level, or FPL. Likewise, to identify the fraction of the population that could be subsidy eligible, I estimate the number of people that fall between 138 percent and 400 percent of FPL. To generate the income distributions, I drew 1,000 random sample for each income group within each county and weighted the results by the adjusted population count.

Endnotes

1   Internal Revenue Service, Statistics of Income, September 19, 2014. Washington, D.C. http://www.irs.gov/uac/SOI-Tax-Stats-County-Data

2   Center for Economic and Policy Research. 2014. March CPS Uniform Extracts, Version 0.9.7 . Washington, DC. http://ceprdata.org/cps-uniform-data-extracts/march-cps-supplement/march-cps-data/

3   United States Census Bureau, Population Estimates, June 26, 2014. Washington, D.C. http://www.census.gov/popest/data/historical/2010s/vintage_2012/county.html

4   Internal Revenue Service, Individual Shared Responsibility Provision – Reporting and Calculating the Payment, March 13, 2015. Washington, D.C. http://www.irs.gov/Affordable-Care-Act/Individuals-and-Families/ACA-Individual-Shared-Responsibility-Provision-Calculating-the-Payment

5   Center for Economic and Policy Research. 2014. March CPS Uniform Extracts, Version 0.9.7 . Washington, DC. http://ceprdata.org/cps-uniform-data-extracts/march-cps-supplement/march-cps-data/

6   Piketty, Thomas, and Emmanuel Saez. *The evolution of top incomes: a historical and international perspective.* No. w11955. National Bureau of Economic Research, 2006. http://eml.berkeley.edu/~saez/piketty-saezAEAPP06.pdf

7   Burkhauser, Richard V., Shuaizhang Feng, Stephen P. Jenkins, and Jeff Larrimore. "Estimating Trends in US Income Inequality Using the Current Population Survey: The Importance of Controlling for Censoring." ISER Working Paper Series 2008, no. 25 (September 2008). http://www.econstor.eu/bitstream/10419/92206/1/2008-25.pdf

8   The Affordable Care Act uses income based on health insurance units that are similar but not necessarily identical to tax units. A health insurance unit is a grouping that can be covered by a health insurance plan and will generally conform to the family which also is similar to a tax unit. For our estimates we will assume tax units and health insurance units to be identical.

Our Mission

Accelerate cutting-edge analysis into whether and how structural changes in the U.S. economy, particularly related to economic inequality, affect economic growth.

Washington Center
for Equitable Growth