Where the Wealth Is: The Geographic Distribution of Wealth in the United States

Rebecca Chenevert, Alfred Gottschalck, Mark Klee, and Xingyou Zhang[1]

Social, Economic and Housing Statistics Division

U.S. Census Bureau

Household net worth, or wealth, is known to exhibit a highly skewed distribution. Estimates of wealth concentration show that the top 0.1 percent of families held 22 percent of the wealth owned by U.S. households in 2012.[2] However, household wealth is a difficult concept to measure. In order to create reliable estimates of net worth for small demographic groups or for subnational geographies, we need a data source that is large enough to allow reliable subgroup analysis and that is comprehensive enough to allow a direct construction of net worth. At present, no such data source exists in the United States.

New research being undertaken at the U.S. Census Bureau combines information from the Survey of Income and Program Participation (SIPP) and the American Community Survey (ACS) to create wealth estimates for smaller geographies and populations than were previously available. The SIPP is nationally representative and has rich and detailed information on wealth, while the ACS has more limited information on wealth but has a very rich sample with a diversity of geographic areas represented. We estimate a model of household net worth on SIPP data and use the resulting estimates to predict net worth for ACS households. This paper presents preliminary estimates of wealth and inequality at sub-national levels from the ACS, and seeks to validate and improve the estimation.

---

[1] The authors wish to thank seminar participants at the Census Bureau and at the 2017 ASSA meetings for valuable input and feedback regarding this work. Contact information is as follows: Chenevert: Rebecca.L.Chenevert@census.gov; Gottschalck: Alfred.O.Gottschalck@census.gov; Klee: Mark.A.Klee@census.gov; Zhang: Xingyou.Zhang@census.gov. This paper is released to inform interested parties of ongoing research and to encourage discussion of work in progress. The views expressed in this paper are those of the authors and not necessarily those of the U.S. Census Bureau. Any errors are our own.

[2] Saez, Emmanuel, and Zucman, Gabriel. 2014. "Wealth Inequality in the United States since 1913: Evidence from Capitalized Tax Data." NBER Working Paper 20625. http://www.nber.org/papers/w20625.

Introduction

Wealth inequality has become an increasing concern of Americans, and many question whether increasing inequality affects economic growth.[3] Indeed, during the recent presidential primary election, debates about inequality took center stage. In 2014, Thomas Piketty's book, Capital in the Twenty-First Century, reached number one on the New York Times bestseller list, so this issue has clearly become a matter of serious public debate. Despite the widespread interest in wealth inequality, the sources of wealth data remain surprisingly limited. This paper seeks to contribute additional data for this discussion by providing the first estimates of wealth for American Community Survey (ACS) households.

When looking for information on inequality, the majority of available information centers on income inequality.[4] While income inequality is also very informative of economic well-being, it can only show the flow of new resources available to the household over a period of a year. By contrast, overall household net worth can show the entire stock of resources available to a household and paint a better portrait of overall well-being. Unfortunately, wealth is a more difficult concept to measure than income. For example, some illiquid assets such as real estate, businesses, and vehicles do not have an observable market value. Tax data provide analysts with a generally high quality measure of income that surveys often mismeasure. However, because only some types of wealth are taxed in the U.S., administrative data tend to miss the value of many types of assets and debts.[5]

Furthermore, the work of Raj Chetty and others (Chetty, Hendren, Kline and Saez, 2014) has shown that *communities* matter for economic outcomes, particularly for children (Chetty, Friedman, and Rockoff, 2014; Chetty, Hendren, and Katz, 2015). What is necessary to study wealth at a community level is a data source that combines survey measures of wealth and a sample size large enough to create comparable estimates across the country. The American Community Survey (ACS) is a natural starting point, as it replaced the Census Long form to create comparable estimates of demographic, social, economic, and housing characteristics across communities. However, the ACS lacks questions about households' wealth. Adding detailed information about wealth would be overly burdensome to the more than two million respondents that answer the ACS every year. To overcome this obstacle, in this paper, we model wealth of ACS households using the relationships evident in the Survey of Income and Program Participation. This allows us to generate estimates of net worth and wealth inequality at the state-level, and opens the door for future work to create estimates at even smaller geographies. In the future, we also plan to use this method to create estimates of wealth for small populations such as immigrants.

The contribution of this work is twofold. First, we provide a new application of small area estimation techniques. We demonstrate in a new setting how to employ a detailed survey like SIPP to create new

[3] See a Aghion, Caroli, and García-Peñalosa (1999).
[4] For example, Kuznets (1955), and Barro (2000) look at the effects of income inequality on growth.
[5] Although some administrative measures of assets do exist for the U.S., they are typically limited in scope. For example, Ameriks et al. (2015) utilize administrative wealth data for accounts held at Vanguard. By contrast, countries such as Denmark that impose a wealth tax do offer more comprehensive administrative wealth data. See Boserup, Kopczuk, and Kreiner (2016) and Fagereng et al. (2016).

data about our nation and economy without adding additional burden to survey respondents. Second, this paper provides estimates of wealth at the state level that have not been available before.

While these results are preliminary, we find that the median value of net worth varies considerably across the nation. We also find that the ratio of wealth to income varies across states, which suggests that residents in some states are better prepared for negative income shocks than others. Finally, we find that the levels of inequality vary across geographies.

1. Data and Methodology

The Survey of Income and Program Participation (SIPP) is the premiere data source for measuring income and participation in government programs. As such, the survey collects an extensive amount of information about the economic situation within households over a collection period of 3-4 years. While the focus of the survey is not on household wealth *per se*, measuring wealth is an important part of measuring household economic well-being. It is also an important component of measuring eligibility for some government programs. As such, the SIPP has a history of measuring detailed components of household net worth and releasing these estimates at the regional level and by demographic group at the national level.

The SIPP has evolved over the years, and has recently been redesigned. In this paper, we use wave 1 of the most recent panel, which was fielded in early 2014 with a 2013 reference period and samples about 27,000 households. In the future, we plan to create a time series by using the data we have available for 2009, 2010, and 2011, as well as data prior to 2009. However, the ACS also had a number of question changes in 2008, so we prefer to employ only data from after these question changes.[6] The SIPP 2014 Panel is currently in production and will have data available annually for 2013-2016. The new design should continue to have wealth data annually on an ongoing basis.

The American Community Survey (ACS) replaced the long form of the Census, and it has been in production since 2005. The primary purpose of the ACS is to provide useful data related to housing, demographics, employment, health insurance and income at the community level, and the sample size is more than two million households. Although it is not designed to study wealth, some key correlates and components of wealth are collected in the ACS. In particular, the questions relating to home value and total household income received in the past year likely serve as important predictors of household wealth

We validate our modeled ACS wealth estimates in part by comparison to the Survey of Consumer Finances (SCF). The SCF, conducted by the Federal Reserve, was designed primarily to measure details of the components of household wealth. As such, its sample design and survey questions are tailored to

[6] We do not currently plan to use the 3-year ACS files from 2009-2011 for two reasons. First, net worth was changing significantly over this time period, so using the 3-year file may be inappropriate even with inflation adjusting. Second, in this paper, we focus on results by state, which we can do with a 1-year file. In future versions, we may use a 3-year or 5-year file in order to look at smaller geographies, but it would be with the caveat that they may be averages across years where wealth is changing rapidly. Also, the 3-year file was discontinued, and so will not be available after the 2011-2013 data years.

collect much more detailed information about wealth at all points in the distribution than other surveys (see Juster, Smith, and Stafford, 1999; Czajka et. al, 2003; Juster and Kuester, 1991; Curtin, Juster, Morgan, 1989; Wolff, 1999). It is commonly assumed that the more detailed questions on the SCF allow for better measurement of the components of net worth, and therefore, a better measure of overall net worth. The estimates of net worth from SIPP have been shown to be lower than those reported in the SCF but the difference has narrowed with the recent redesigns in SIPP data collection (Czajka et. al., 2003; Eggleston and Klee, 2015; Eggleston and Gideon, forthcoming; Eggleston and Reeder, forthcoming). Wave 1 of the 2014 SIPP Panel seems to perform well relative to the SCF across much of the distribution, and it offers several advantages over other datasets for modeling wealth. First, detailed geography information is available to be joined with the ACS. Second, in the future we hope to link various administrative data sources to both SIPP and ACS to assist in the modeling.

Both SIPP and ACS were geocoded to census block level, which allowed us to link both individual-level but also area-level factors for model fitting and prediction in a multilevel modeling framework. Thus, we decided to use a Multilevel Regression and Poststratification (MRP) model to generate state-level small area estimates of household net worth (see details on MRP in the Appendix A.2). The basic idea of MRP approach for small area estimation is 1) to construct and fit the multilevel model for the outcome of interest and 2) then make predictions of the outcome using a large survey or census population data to produce its reliable small area estimates. In this study, we first used SIPP data to construct a multilevel linear regression model for household net worth (equation 1) and then made prediction with ACS, the largest demographic survey that could produce reliable state-level survey estimates (equation 2).

$$(1)\ \ Y_{is}^{SIPP} = x_{is}^{SIPP}\beta + \theta_s + \varepsilon_{is}$$

In equation (1), Y is the vector of modeling outcome: the rank of household net wealth in SIPP; x is the matrix of covariates (fixed effects), including householder's age, sex, race/ethnicity, marital status, education, disability status, household property values and total income, residential place urban/rural status, residential census tract medium household value and income; β is the vector of their corresponding regression coefficients; θ the state-level random effects; and ε is the residual random effects; both are assumed to be independent and normally distributed with a mean of zero. Here the outcome used in the multilevel linear model is the rank of SIPP household net worth, because the original SIPP household net worth was very skewed and heavily tailed (the weighted skewness is 41 and kurtosis is 2027; the unweighted skewness is 31 and kurtosis is 1784).

The above multilevel linear model was fitted in SAS using proc GLIMMIX and applied to ACS household data to predict individual-level household net worth as in equation (2):

$$(2)\ \ \widehat{Y_{is}}^{ACS} = X_{is}^{ACS}\hat{\beta} + \widehat{\theta_s}$$

where $\hat{Y}$ is the vector of predicted net worth ranks for ACS households; X is the matrix of covariates from ACS households; $\hat{\beta}$ is the vector of their estimated regression coefficients; $\hat{\theta}$ the estimated state-level random effects.

To obtain the quantities of our interest, the predicted net worth ranks of ACS households were converted to the net worth values in terms of U.S. dollars. As expected, the predicted rank could be less than one (the richest household in SIPP) or negative, and also could be larger than the maximum rank (the poorest in SIPP). This means ACS households could have higher or lower net worth ranks than those in SIPP, and therefore have a larger range of household net worth values. Although we know that the tails of the wealth distribution are skewed, we used a linear extrapolation method to calculate net worth values for these cases. This should be conservative in the sense that the tails should both have estimates slightly biased toward the mean. Because this is a small part of the distribution, it should not have much of an effect on our overall estimates of median net worth or the points along the distribution that we estimate, but it could have a small downward bias on measures of the Gini coefficient.

We treated these predicted values as known in ACS and generated final state-level estimates of household net worth values (median and mean). In order to create standard errors for the median net worth estimates, we accounted for the uncertainty in both model prediction and the variations in ACS sampling via Monte Carlo simulation. We do this by using the standard errors of the predicted model coefficients and the random effects to create 1000 simulations, which accounts for modeling uncertainty. Then sampling variation from the ACS is accounted for using the Balanced Repeated Replication (BRR) method, using one draw from each of the 1000 Monte Carlo simulations.

Table 1 shows the estimated regression coefficients. The dependent variable is the rank in the wealth distribution, so negative coefficients correspond with higher wealth households. In general, we find expected patterns. Households where the householder is older, more educated, married, and with a more valuable home are wealthier.[7] We find that higher income households are wealthier up until an annual income of about $200,000. Households where the householder is disabled or those in urban areas tend to be less wealthy.[8] Neighborhood characteristics (defined by the Census tract medians for 2008 through 2012) show that higher income and property values of neighbors are also associated with higher wealth.

While direct interpretation of these coefficients is difficult, we do still prefer to do a model with rank as the outcome variable. Using the rank as the outcome variable accounts for there being differential effects at different points in the distribution. As an example, the coefficient on the non-Hispanic white

---

[7] The sampling frames of the SIPP and ACS are different. Nevertheless, our results are comparable regardless of whether we apply survey weights. This likely results because we control for many of the demographic and neighborhood characteristics which are used to construct survey weights. The results presented here do not apply survey weights. The householder is a person who owns or rents the home. Because married couples could have either spouse be designated as the householder, we included demographics of both spouses in earlier versions. However, values tend to be correlated and most of these were dropped due to insignificance (with the exception of the sex and marital status interaction variables).

[8] In the SIPP, the unit of observation for net worth is the household. While one might argue that the co-residing family unit is a preferable unit of observation, few estimates of the civilian population include those in group quarters housing. Although these individuals are included in the ACS sampling frame, and some types of group quarters are available in the SIPP, we do not estimate net worth for those in group quarters housing. The SCF, discussed in more detail in Section 4, uses a family as the unit of observation. However, since both the SIPP and ACS sample a household, we have chosen to use the household as well. Further work could test the extent to which this difference in units of observation drives differences in SCF and modeled ACS estimates of net worth.

indicator is roughly -1000, compared with the reference group of Hispanics. This means that *ceteris paribus*, the average difference between non-Hispanic Whites and Hispanics along the distribution of about 27,000 households is about 1000. The concentration of these two demographic groups is different along the distribution, and so using a ranking model helps account for that. For that reason, we find a ranking model preferable to using a percentile model. We have considered using a model that ranks only unique values. This would make an impact for values such as zero, where there is a mass of households located.

Table 1.

Regression of Net Worth Rank on Household Chacteristics

Effect	Estimate	Standard Error	P value
Intercept	10011.0	850.7	<.0001
Male	-46.7	93.1	0.6160
Age 15-24	8724.3	1313.6	<.0001
25-29	10019.0	818.7	<.0001
30-34	9545.3	788.4	<.0001
35-39	8384.5	789.2	<.0001
40-44	7313.5	784.4	<.0001
45-49	5656.1	790.0	<.0001
50-54	5466.7	784.9	<.0001
55-59	3989.7	784.5	<.0001
60-64	3231.0	774.3	<.0001
65-69	2753.6	781.0	0.0004
70-74	2003.1	815.2	0.0140
75-79	2026.0	857.6	0.0182
80-84	1588.0	913.9	0.0823
85+ (ref)	0.0	.	.
Non-Hispanic White	-947.1	112.1	<.0001
Non-Hispanic Black	275.5	132.9	0.0382
Non-Hispanic Asian	-1075.0	194.3	<.0001
Non-Hispanic Other races	259.6	235.5	0.2702
Hispanic (ref)	0.0	.	.
Less high school	4002.3	805.8	<.0001
High school	1973.6	779.5	0.0114
Some college	1344.1	839.4	0.1093
College	1279.3	897.3	0.1540
Graduate Degree (ref)	0.0	.	.
Never Married	296.4	130.9	0.0235
Previously married	990.7	106.3	<.0001
Married with spouse present (ref)	0.0	.	.

continued...

Table 1. **Regression of Net Worth Rank on Household Chacteristics *(continued)***			
Effect	Estimate	Standard Error	P value
Disability	1165.7	81.4	<.0001
Urban	709.0	85.6	<.0001
Home Owner	-5625.1	84.8	<.0001
Household Income (dollars) <=0	4133.8	453.5	<.0001
<5,000	3640.2	430.5	<.0001
<15,000	4338.0	403.4	<.0001
<20,000	4158.9	410.1	<.0001
<25,000	3809.9	409.4	<.0001
<30,000	3347.7	409.5	<.0001
<35,000	3677.4	410.4	<.0001
<45,000	2999.4	399.0	<.0001
<55,000	2720.2	400.4	<.0001
<65,000	2555.5	402.8	<.0001
<75,000	1919.6	405.3	<.0001
<90,000	1571.9	401.3	<.0001
<105,000	947.6	405.5	0.0194
<125,000	228.7	407.4	0.5745
<150,000	-204.7	412.4	0.6196
<200,000	-988.8	409.0	0.0156
<500,000	-1938.3	412.4	<.0001
>=500,000 (ref)	0.0	.	.
Residental Property Value (per $10,000)	-37.4	1.4	<.0001
Sex x marital status	Yes		
Age x education	Yes		
Tract Median Household Income (per $10,000)	-174.0	18.9	<.0001
Tract Median Property Value (per $10,000)	-33.7	3.5	<.0001
Source: Survey of Income and Program Participation, 2014 Panel, Wave 1.			

Table 2: Compare the SIPP and ACS Samples.

Cannot display comparisons until SIPP data are released.

2. Results

Overall, we find considerable variation across states in the median value of net worth, as well as differences across the distribution. The fact that we find variation is not surprising, but it is encouraging that we find some expected patterns.

Table 3.
Net Worth by State

	Median Net Worth	Mean Net Worth	Median Home Value
United States	**96,679**	**2,561,365**	**173,900**
Alabama	83,349	621,898	122,700
Alaska	120,365	988,504	254,000
Arizona	79,785	1,371,127	166,000
Arkansas	78,554	446,046	109,500
California	96,190	8,452,846	373,100
Colorado	118,180	2,381,632	240,500
Connecticut	147,278	6,718,527	267,000
Delaware	126,219	1,728,596	226,200
District of Columbia	52,201	11,153,890	470,500
Florida	88,938	1,760,407	153,300
Georgia	78,710	956,223	141,600
Hawaii	153,570	8,070,434	500,000
Idaho	95,389	761,704	159,000
Illinois	102,768	1,612,807	169,600
Indiana	90,247	431,133	122,200
Iowa	108,512	407,859	126,900
Kansas	96,608	564,284	129,700
Kentucky	87,998	551,148	120,900
Louisiana	86,574	588,047	140,300
Maine	115,971	763,743	172,800
Maryland	136,853	4,882,869	280,200
Massachusetts	148,838	4,690,209	327,200
Michigan	87,983	555,774	117,500
Minnesota	133,224	851,740	180,100
Mississippi	75,772	511,858	97,500
Missouri	91,123	630,802	133,200
Montana	112,580	1,210,327	190,100
Nebraska	96,347	381,617	132,700
Nevada	63,224	1,181,840	165,300
New Hampshire	148,468	1,461,193	233,300
New Jersey	143,831	4,257,367	307,700
New Mexico	88,135	1,268,174	159,200
New York	100,543	5,584,790	277,600
North Carolina	93,956	1,116,113	154,300
North Dakota	103,615	438,364	155,400
Ohio	87,717	418,669	127,000
Oklahoma	82,256	380,345	116,500
Oregon	93,621	815,600	229,700
Pennsylvania	113,131	1,073,438	164,200
Rhode Island	108,967	1,917,104	232,300
South Carolina	93,925	1,234,139	139,200
South Dakota	99,726	804,990	138,400
Tennessee	87,508	776,648	140,300
Texas	78,825	1,220,117	132,000
Utah	104,950	1,104,458	211,400
Vermont	141,716	1,367,707	218,300
Virginia	119,459	4,240,759	239,300
Washington	106,626	1,800,999	250,800
West Virginia	92,262	401,557	103,200
Wisconsin	111,986	594,762	163,000
Wyoming	119,763	1,178,441	195,500

Source: Columns (1) and (2), modeled estimates of net worth from the American Community Survey, 2013. Column (3), American Community Survey, 2013.

As expected, states in the Mid-Atlantic, Northeast, and Hawaii are among the wealthiest states at the median. Also as expected, states in the Southeast are among the poorest states at the median. Although the District of Columbia is the poorest state at the median, wealth inequality in the District of Columbia is among the highest in the nation according to the Gini coefficient. Wealth inequality is also relatively high in the Mid-Atlantic, Northeast, and Pacific Coast.

Table 4.

Measures of Inequality by State

					Median Net Worth by Quintile				
	Gini Coefficient	90/10 Wealth Ratio	90/50 Wealth Ratio	50/10 Wealth Ratio	First Quinitile	Second Quinitile	Third Quintile	Fourth Quintile	Fifth Quintile
United States	**0.97**	**292.1**	**3.8**	**77.4**	**1250**	**27914**	**96739**	**173829**	**365139**
Alabama	0.92	1297.4	3.1	417.9	200	26,701	83,559	141,889	259,460
Alaska	0.92	117.7	3.3	35.3	3,415	36,317	120,581	212,544	400,767
Arizona	0.96	1528.3	3.8	401.2	199	16,754	79,943	153,296	304,536
Arkansas	0.89	1176.2	3.0	394.1	198	19,433	78,222	133,276	233,538
California	0.96	329.7	7.5	44.2	2,174	22,246	96,216	221,012	716,830
Colorado	0.96	206.0	3.8	54.7	2,164	34,914	118,287	212,542	445,693
Connecticut	0.96	311.8	4.1	75.2	1,954	47,982	147,244	254,226	610,189
Delaware	0.95	149.0	3.0	50.2	2,525	56,858	126,661	203,170	376,158
District of Columbia	0.95	756.6	16.8	44.9	1,154	11,312	52,005	181,283	875,651
Florida	0.96	572.8	3.4	169.0	527	23,552	89,092	156,029	301,861
Georgia	0.95	1456.3	3.7	395.3	200	14,920	78,961	144,452	290,923
Hawaii	0.95	153.8	6.1	25.1	6,110	43,725	153,391	322,168	939,412
Idaho	0.92	222.8	2.9	76.5	1,249	37,094	95,595	159,095	278,302
Illinois	0.95	263.7	3.4	77.8	1,325	32,748	103,133	177,441	349,304
Indiana	0.88	498.9	2.8	181.3	498	30,155	90,241	147,754	248,429
Iowa	0.83	149.8	2.6	57.0	1,908	49,236	108,705	170,002	285,783
Kansas	0.89	285.9	3.1	91.9	1,050	30,796	96,516	166,176	300,154
Kentucky	0.90	244.8	2.9	83.8	1,050	28,550	87,990	144,486	257,056
Louisiana	0.91	496.1	3.1	160.5	538	23,868	86,392	147,082	267,092
Maine	0.90	116.1	2.8	41.3	2,826	54,262	116,794	180,721	328,602
Maryland	0.96	227.1	4.3	52.3	2,614	48,620	136,803	245,197	593,564
Massachusetts	0.96	179.9	4.1	43.8	3,399	39,255	148,942	264,318	611,573
Michigan	0.91	2017.7	2.9	697.4	126	30,912	88,051	145,572	254,735
Minnesota	0.90	115.2	2.8	41.7	3,194	63,431	133,276	205,200	367,831
Mississippi	0.91	1546.7	3.1	507.1	149	20,977	75,750	126,450	231,124
Missouri	0.91	301.9	3.0	99.7	916	28,002	91,333	153,392	276,427
Montana	0.94	143.5	2.9	49.2	2,288	40,065	112,379	182,908	328,403
Nebraska	0.84	275.9	2.9	94.7	1,020	30,199	96,595	161,919	281,400
Nevada	0.96	551.8	4.4	124.6	505	11,224	62,986	139,197	278,991
New Hampshire	0.93	95.4	2.9	32.7	4,539	66,079	148,330	232,376	432,813
New Jersey	0.96	229.9	4.3	53.5	2,700	41,953	144,392	254,984	620,583
New Mexico	0.95	310.6	3.5	88.4	999	33,352	88,359	154,508	310,377
New York	0.97	189.2	4.8	39.4	2,548	24,999	100,485	200,781	482,086
North Carolina	0.94	263.5	3.4	78.5	1,201	23,551	94,256	163,680	316,469
North Dakota	0.86	153.6	2.8	55.6	1,877	33,706	104,382	167,577	288,361
Ohio	0.88	2220.2	3.0	748.3	117	22,573	87,765	147,854	260,387
Oklahoma	0.87	509.8	3.1	164.7	500	20,602	82,278	142,417	254,698
Oregon	0.93	320.2	3.4	94.2	992	18,400	93,463	169,779	317,692
Pennsylvania	0.93	171.9	3.0	57.4	1,969	45,707	113,270	181,477	339,352
Rhode Island	0.96	294.7	3.7	79.6	1,373	21,988	109,284	201,069	404,637
South Carolina	0.95	263.6	3.2	81.5	1,156	34,096	94,299	158,475	305,003
South Dakota	0.93	199.4	2.8	71.7	1,399	36,466	100,301	162,617	279,131
Tennessee	0.93	503.5	3.1	160.5	546	24,197	87,578	146,163	274,749
Texas	0.95	457.4	4.0	113.8	694	17,828	78,960	151,965	317,274
Utah	0.94	214.0	3.2	67.1	1,565	44,304	105,099	178,633	335,150
Vermont	0.93	96.1	2.7	35.0	4,035	65,004	141,478	218,613	387,979
Virginia	0.96	237.1	4.6	51.7	2,310	43,747	119,348	217,346	547,584
Washington	0.96	265.1	3.6	72.7	1,470	26,147	106,930	194,207	389,717
West Virginia	0.86	155.1	2.5	62.8	1,467	43,004	92,072	141,888	227,554
Wisconsin	0.89	181.3	2.7	66.7	1,682	40,396	112,205	177,820	305,201
Wyoming	0.93	118.6	2.7	43.9	2,740	47,347	120,562	186,728	325,094

Source: Modeled estimates of net worth from the American Community Survey, 2013

The compression in the middle of the distribution in the modeled estimates lead to very high measures of the Gini coefficient, presented in Table 4.[9] Most likely, the best solution to this unpalatable implication of our modeled estimates would be to modify the model. One possible modification is to model net worth using some specified distribution rather than modeling net worth rank. Another possible modification is to extrapolate the net worth of the wealthiest and poorest households by employing a higher order spline rather than a linear spline.

3. Validation

As estimates of net worth by state are being newly released, we do not have many other sources to compare with in order to evaluate quality. However, we expect that certain patterns will hold and we can verify whether we see these patterns in the data. For example, in Table 3 we included median home values from the ACS as a reference, since the typical household holds much of their worth in the value of their home (Gottschalck, 2008). These values are either reported by ACS respondents or hot-deck imputed. Since these estimates are not generated by our model, we can evaluate the quality of our model by comparing patterns in median net worth by state implied by our model to patterns in median home values by state that are already present in ACS data.

Further, we can evaluate how well the national values of net worth implied by our model correspond to the distribution from several nationally representative surveys of wealth. In Table 5, we show the distribution of net worth from the SCF compared with that of the modeled ACS. We also present in Table 6 the income Gini coefficients by state (Noss, 2014) from the American Community Survey. While we expect the levels of the Gini coefficients to be different by state, we expect that states with higher income inequality would also likely have higher wealth inequality.

The overall level of median net worth is higher in the modeled estimates than what we see in other survey estimates of the nation as a whole. In particular, the modeled ACS estimate is about $96,700 compared with $81,000 in the SCF. In general, the modeled distribution seems to be more condensed than the distribution suggested by the SCF. The modeled ACS estimate of net worth is higher at the 20th percentile of the distribution and lower at the 80th percentile of the distribution relative to the SCF. The modeled ACS estimate is substantially lower than the SCF estimate at the 90th percentile. In principle, estimates from the SIPP or the ACS should be slightly higher than those from the SCF in some parts of the distribution because the unit of observation is a household, rather than a co-residing family.

[9] The Gini coefficient was calculated numerically using the method outlined by Phillip Cohen. We used the ACS modeled estimates for each household and the associated survey weights.

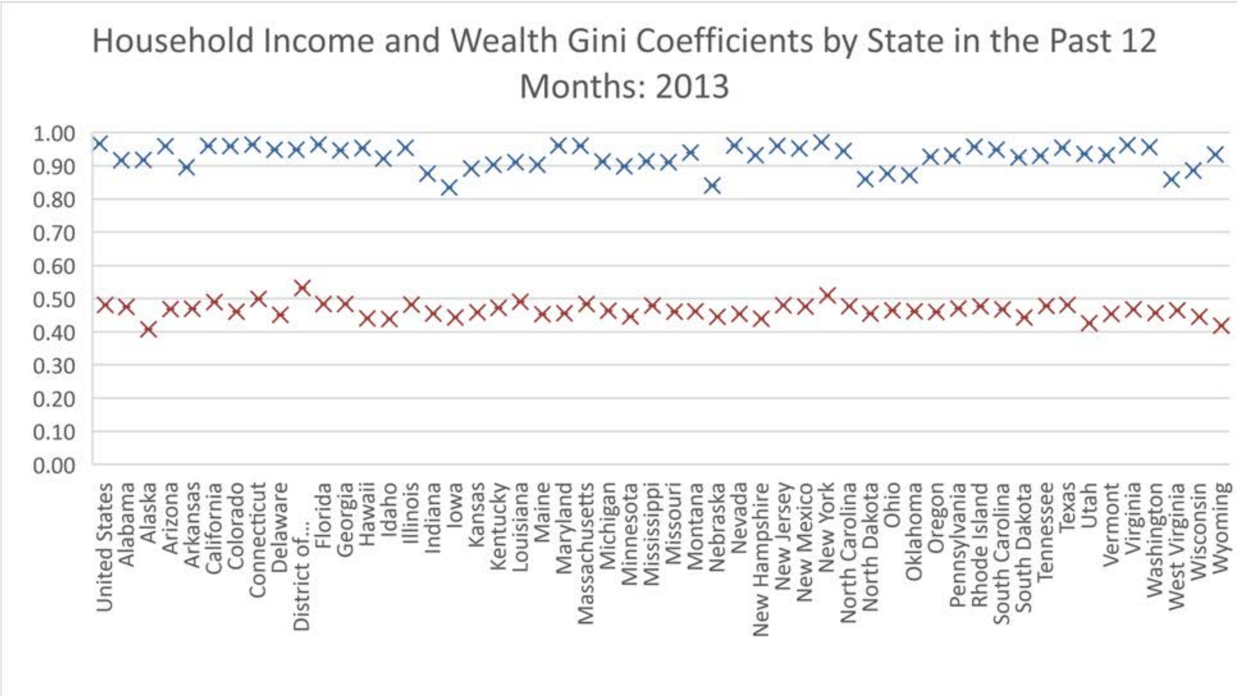
Table 5.
National Distribution of Net Worth from Modeled and Survey Estimates

Percentile	ACS Modeled Estimates	SCF Estimates
10	1,250	-2,072
20	6,826	4,261
30	27,921	14,739
40	64,085	38,043
50	96,740	81,049
60	132,611	146,952
70	173,829	245,750
80	233,532	425,809
90	365,139	939,123

Source: Source: Column (1), modeled estimates of net worth from the American Community Survey, 2013. Column (2), Author's calculations from the Survey of Consumer Finances, 2013.

While the actual values of the Gini coefficients are also high relative to what appears in other national surveys of wealth, we do see some patterns that look similar between income inequality as measured from direct survey estimates of the ACS and wealth inequality from the modeled estimates. While the model needs refinement to produce better estimates of the wealth distribution, the pattern across states is encouraging. See Figure 1 below, or Table 6 in the appendix.

Figure 1.



4. Planned Future Work

What we present here is a first step in a much larger project. As we move forward in this work, we hope to link SIPP and ACS to additional data sources that will help provide better estimates of household net worth. In particular, we plan to seek approval to link individuals in the SIPP and ACS to tax records such as the Detailed Earnings Record (DER), Internal Revenue Service (IRS) 1040 data, and IRS 1099 data. The DER is provided from the Social Security Administration (SSA) and has the complete earnings history of individuals going back to 1978. Prior to 1978, the Summary Earnings Record (SER) also has information on earnings which dates back to the 1950s, but these earnings are capped at the social security taxable maximum. This information could be useful, as many who are retired or close to retirement likely have wealth more closely correlated with their earnings history than with current earnings.

Information from the IRS 1040 returns could also help better predict wealth from income earning assets. While survey respondents are asked to report income from all sources on both the SIPP and the ACS, less salient income such as stock dividends that are reinvested or small amounts of interest income are likely to be under-reported (Eggleston and Reeder, 2016). If receipt of these types of income is not reported, and particularly if it is reported differentially in the two surveys, the survey measures of asset income that we currently use to predict wealth are likely less well suited as a predictor of net worth than administrative measures of asset income. Indeed, we have reason to suspect that asset income is better reported in the SIPP than in the ACS. SIPP includes a battery of questions about income from each of a large set of asset types. By contrast, ACS asks respondents to report total income from all assets combined. This asymmetry across surveys might yield differential reporting for less salient assets. Adding information from 1040 returns can help to put the two surveys on equal footing, and relationships between income and reported values in the SIPP can help to project actual values for the ACS respondents.

Finally, we plan to expand this work by including additional years of data. As was mentioned earlier, the SIPP underwent a redesign and many improvements were made to obtain better quality wealth data. While the 2014 Panel data may not be comparable in all ways to the 2008 Panel, we may still find useful relationships in a time series. Further, to allow analysts to compute time trends of wealth by subnational geography or small demographic group, we could extend this work back to earlier years. In particular, we would like to include the 1990 and 2000 Census long form data and the 1990 and 1996/2001 SIPP panels to estimate wealth from those time periods.

Ultimately, the methodology we apply in this paper has many potential applications across various socioeconomic outcomes. We plan to use the results from these other applications to aid us in refining the above wealth estimates.

5. Conclusion

Measuring wealth provides information about overall household well-being, and can inform us about the total resources available to a household. Measuring net worth is very difficult because of the many types of assets and liabilities held by a household as well as the sensitive nature of survey questions about wealth. Also, because wealth itself is not taxed, income tax data can only paint part of the picture. This project serves as a first step in trying to estimate the wealth of households by state. While results are preliminary, we find that the median value of net worth varies considerably across the nation. We also find that measures of inequality and the wealth distribution are very different across the states, and the wealth distribution differs from the income distribution.

In the future, we plan to further investigate alternative models for estimating wealth for smaller geographies. While we are capturing variation across the states, the levels are more compressed than what we would expect given other sources of wealth data at the national level.

References

Aghion, Philippe, Eve Caroli and Cecilia García-Peñalosa. "Inequality and Economic Growth: The Perspective of the New Growth Theories." *Journal of Economic Literature,* Vol. 37, No. 4, December 1999. https://dash.harvard.edu/handle/1/12502063. Ameriks, John, Andrew Caplin, Minjoon Lee, Matthew D. Shapiro, and Christopher Tonetti. "The Wealth of Wealthholders." NBER Working Paper 20972, February 2015. http://www.nber.org/papers/w20972.pdf.

Boserup, Simon H, Wojciech Kopczuk, and Claus T. Kreiner.. "Born with a Silver Spoon? Danish Evidence on Wealth Inequality in Childhood." Working Paper, August 2016. http://www.columbia.edu/wk2110/bin/KidsWealth.pdf.

Barro, Robert J. "Inequality and Growth in a Panel of Countries *Journal of Economic Growth."* Vol. 5, No. 1, February 2000. http://scholar.harvard.edu/files/barro/files/inequality-growth-1999.pdf.

Chetty, Raj, John N. Friedman, and Jonah E.Rockoff. "Measuring the Impact of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood." *American Economic Review,* Vol. 104, No. 9, September 2014. http://www.ingentaconnect.com/content/aea/aer/2014/00000104/00000009/art00002.

Chetty, Raj, Nathaniel Hendren, and Lawrence F. Katz. "The Effects of Exposure to Better Neighborhoods on Children: New Evidence from the Moving to Opportunity Experiment." NBER Working Paper 21156, May 2015. http://www.nber.org/papers/w21156.

Chetty, Raj, Nathaniel Hendren, Patrick Kline, and Emmanuel Saez. "Where is the Land of Opportunity? The Geography of Intergenerational Mobility in the United States." NBER Working Paper 19843, January 2014. http://www.nber.org/papers/w19843.

Cohen, Philip. Gini Code. (SAS Code for calculating Gini coefficients.) https://www.terpconnect.umd.edu/~pnc/gini.sas.

Czajka, John, Jonathan E. Jacobson, and Scott Cody. 2003. "Survey Estimates of Wealth: A Comparative Analysis and Review of the Survey of Income and Program Participation." Mathematica Policy Research Report PR03-45, Washington, DC. https://www.mathematica-mpr.com/-/media/publications/pdfs/estimateswealth.pdf.

Eggleston, Jonathan and Klee, Mark. "Reassessing Wealth Data Quality in the Survey of Income and Program Participation." SEHSD Working Paper number SEHSD-WP2016-17, February 2015. https://www.census.gov/content/dam/Census/library/working-papers/2016/demo/SEHSD-WP2016-17.pdf

Eggleston, Jonathan and Gideon, Michael. "Evaluating Wealth Data Quality in the 2014 Panel of the Survey of Income and Program Participation." Forthcoming.

Eggleston, Jonathan and Reeder, Lori. "Does Encouraging Record Use for Financial Assets Improve Data Accuracy?" Forthcoming.

Fagereng, Andreas, Luigi Guiso, Davide Malacrino, and Luigi Pistaferri. "Heterogeneity in Returns to Wealth and the Measurement of Wealth Inequality." *American Economic Review: Papers and*

Proceedings, Vol. 106, No.5, May 2016 https://www.aeaweb.org/articles?id=10.1257/aer.p20161022.

Gelman A, Little RJA. 1997. "Poststratification into many categories using hierarchical logistic regression." *Survey Methodologist* 23: 127-135.

Gottschalck, Alfred O. "Net Worth and Asset Ownership of Households: 2002." *Current Population Reports, P70-115.* U.S. Census Bureau, Washington, DC., April 2008. https://www.census.gov/prod/2008pubs/p70-115.pdf.

Juster, F. Thomas, and Kathleen A. Kuester. "Differences in the Measurement of Wealth, Wealth Inequality, and Wealth Composition Obtained from Alternative U.S. Wealth Surveys." *Review of Income and Wealth,* Series 37 No. 1, March 1991. http://roiw.org/1991/33.pdf.

Juster, F. Thomas, James P. Smith and Frank Stafford. 1999. "The Measurement and Structure of Household Wealth." *Labour Economics*, Vol. 6. http://www.sciencedirect.com/science/article/pii/S0927537199000123.

Kuznets, Simon. "Economic Growth and Income Inequality," *The American Economic Review*, Vol. 45, No. 1, March 1955. http://assets.aeaweb.org/assets/production/journals/aer/top20/45.1.1-28.pdf.

Little RJA. 1993. "Post Stratification: A Modeler's Perspective." *Journal of the American Statistical Association*, 88:1001-1012.

Noss, Amanda. "Household Income: 2013." *American Community Survey Briefs,* September 2014. https://www.census.gov/content/dam/Census/library/publications/2014/acs/acsbr13-02.pdf.

Park DK, Gelman A, Bafumi J. 2004. "Bayesian Multilevel Estimation with Poststratification: State-level Estimates from National Polls." *Political Analysis* 53(1): 107-121.

Picketty, Thomas. *Capital in the Twenty-First Century.* Harvard University Press. 2014. https://dowbor.org/blog/wp-content/uploads/2014/06/14Thomas-Piketty.pdf.

Saez, Emmanuel, and Zucman, Gabriel. "Wealth Inequality in the United States since 1913: Evidence from Capitalized Tax Data." NBER Working Paper 20625, October 2014. http://www.nber.org/papers/w20625.

Tanenhaus, Sam. "Hey, Big Thinker." *The New York Times*, April 25, 2014. https://www.nytimes.com/2014/04/27/fashion/Thomas-Piketty-the-Economist-Behind-Capital-in-the-Twenty-First-Century-sensation.html?_r=0.

Zhang X, Holt JB, Lu H, Wheaton AG, Ford ES, Greenlund K, Croft JB. 2014. "Multilevel regression and poststratification for small-area estimation of population health outcomes: a case study of chronic obstructive pulmonary disease prevalence using the Behavioral Risk Factor Surveillance System." *American Journal of Epidemiology,* 179(8):1025–1033.

Appendix A.1

Table 6.
Median Household Income and Gini Index in the Past 12 Months by State: 2013

	2013 ACS Median Income (dollars)		2013 ACS Income Gini Coefficient	
	Estimate	MOE	Estimate	MOE
United States	52,250	65	0.481	0.001
Alabama	42,849	641	0.475	0.004
Alaska	72,237	1,892	0.408	0.01
Arizona	48,510	587	0.468	0.005
Arkansas	40,511	710	0.469	0.008
California	60,190	255	0.49	0.002
Colorado	58,823	808	0.461	0.004
Connecticut	67,098	1,058	0.499	0.005
Delaware	57,846	1,876	0.451	0.01
District of Columbia	67,572	3,383	0.532	0.012
Florida	46,036	310	0.484	0.003
Georgia	47,829	628	0.484	0.004
Hawaii	68,020	1,523	0.44	0.009
Idaho	46,783	930	0.438	0.008
Illinois	56,210	403	0.482	0.003
Indiana	47,529	516	0.455	0.005
Iowa	52,229	533	0.443	0.005
Kansas	50,972	609	0.459	0.009
Kentucky	43,399	650	0.472	0.007
Louisiana	44,164	869	0.491	0.005
Maine	46,974	797	0.453	0.007
Maryland	72,483	718	0.456	0.004
Massachusetts	66,768	715	0.484	0.004
Michigan	48,273	378	0.464	0.003
Minnesota	60,702	432	0.446	0.004
Mississippi	37,963	1,029	0.479	0.006
Missouri	46,931	427	0.461	0.004
Montana	46,972	1,140	0.462	0.009
Nebraska	51,440	493	0.445	0.006
Nevada	51,230	589	0.454	0.008
New Hampshire	64,230	1,347	0.439	0.009
New Jersey	70,165	546	0.48	0.003
New Mexico	43,872	950	0.476	0.006
New York	57,369	431	0.51	0.004
North Carolina	45,906	424	0.477	0.004
North Dakota	55,759	1,452	0.455	0.009
Ohio	48,081	406	0.465	0.003
Oklahoma	45,690	534	0.462	0.005
Oregon	50,251	532	0.46	0.006
Pennsylvania	52,007	256	0.47	0.003
Rhode Island	55,902	1,902	0.477	0.011
South Carolina	44,163	659	0.467	0.004
South Dakota	48,947	1,091	0.443	0.009
Tennessee	44,297	501	0.478	0.004
Texas	51,704	238	0.481	0.003
Utah	59,770	762	0.426	0.006
Vermont	52,578	1,561	0.454	0.013
Virginia	62,666	665	0.467	0.003
Washington	58,405	671	0.457	0.004
West Virginia	41,253	746	0.465	0.007
Wisconsin	51,467	370	0.445	0.003
Wyoming	58,752	1,796	0.418	0.012

Source: U.S. Census Bureau, 2013 American Community Survey. Reprinted from Noss (2014)

Appendix A.2

Background

Various small area estimation methods have been developed to generate reliable small area estimates for socioeconomic outcomes of interest using the national surveys. For model-based approach of small area estimation, there are two basic small area models: area-level model and unit-level model. Area-level models model the aggregated outcomes from the original measures of survey respondents while unit-level models use the original measures of survey respondents as model outcomes. The selection and choice of small area estimation models depends on a number of factors, including but not limited to the availability of survey microdata as well as population data. At U.S. Census Bureau, the Survey of Income and Program Participation (SIPP) and American Community Survey (ACS) are both geocoded to census block level, which allows us to link both SIPP and ACS individual-level household characteristics with their area-level factors at multiple geographic levels, such as census tract, county and state. Thus, we employed a specific statistical modeling technique, multilevel regression and poststratification (MRP), for small area estimation using geocoded SIPP and ACS.

The original idea of MRP, especially the poststratification concept, has been discussed for small sample inference by Little (1993). The role of poststratification is to reduce variance in final population estimates and could also reduce bias introduced by survey nonresponse or sampling frame errors. Gelman et al. (1997) further extended this idea and combined multilevel regression models and poststratification to estimate quantities for small domains using survey data. MRP takes three basic steps:

1) construct and fit a multilevel regression model to quantify the relationship between individual survey responses (outcome of interest) and individual demographics and geographic (area-level) factors;
2) apply the fitted multilevel regression model to predict the estimates for each specific geodemographic group;
3) apply poststratification to generate final small area estimates of interest in which the estimates for specific geodemographic groups are aggregated weighted by their corresponding population counts.

The MRP approach has been used for small area estimation in political science (Park et al. 2005) and public health (Zhang et al. 2015). The popular MRP approach usually uses the census population data for prediction, thus could only include very limited individual level variables in the model, such as age, gender, race/ethnicity. In order to overcome this limit, a larger survey can be used for prediction to produce direct reliable estimates for the geographic level of interest. For example, Jerry (2017) combined SIPP and ACS to estimate state-level disability prevalence. Thus, in this paper, we used SIPP for model construction and fitting and used ACS for model prediction.

Statistical Modeling Framework

The revised MRP approach uses SIPP take similar three basic steps: 1) construct and fit a multilevel linear model with SIPP data; 2) make prediction with ACS data; and 3) generate small area estimates using the ACS with the predicted outcome of interest.

1) Model construction and fitting with SIPP data

SIPP is used to fit an appropriate econometric model to estimate household wealth as equation 1. The potential factors included but not limited to household owner age, gender, race, education, the number of household size, and income. Since the geocoded SIPP and ACS data both have census block identifiers, census tract-level, county-level and state-level contextual effects could be introduced into the unit-level multilevel model. This multilevel model takes the generalized linear mixed model form:

$Y_{is} = X\beta + S + \varepsilon$ (1)

Y_{is} is the household economic wealth for household(i) in state(*s*).

X is the vector or list of predictors or covariates or factors associated with economic wealth, including individual householder age, gender, race/ethnicity, education, as well as census tract, county and state level contextual factors, such as county rural/urban status.

S is the state-level random effect.

ε is the residual random effect or also called random error.

All state and residual random effects are assumed to statistically independent from each other and follow normal distribution. The above multilevel model could be fitted using Proc GLIMMIX or MIXED in SAS.

2) Model prediction with large survey data (ACS)

Unit-level multilevel model prediction requires that all the predictors must be available for all the units in the target population of interest. This requires all the variables included in the multilevel model based on SIPP should also be available for the large survey, ACS. Thus, the expected household wealth could be predicted from the fitted multilevel linear model for each household in ACS (equation 2).

$E(\hat{Y}_{is}) = X\hat{\beta} + \hat{S}$ (2)

$\hat{Y}_{is}$ is the predicted household economic wealth for all sampled ACS households. With the predicted values for each individual household sample in ACS, we could follow the classic survey estimation process to generate estimates of interest.

3) Small Area Estimates with ACS

In this study, we are only interested in state-level estimates for household wealth outcomes. With single year ACS for prediction, we could generate reliable estimates for states. However, there are two types of uncertainties we should consider in our final state-level estimates of household wealth outcomes: model uncertainties and ACS sample variability. In other words, we need estimate the mean square errors (MSEs) associated with each small area estimates. We apply Monte Carlo simulation technique to accomplish this objective. It takes three steps:

1) Make 1,000 predictions for each household in the ACS sample; technically we simulate 1,000 β_b where $\beta_b \sim N(\hat{\beta}, \delta_{\beta}^2)$; similarly we simulate 1,000 $S_b$where $S_b \sim N(\hat{S}, \delta_s^2)$; δ_{β} and δ_s are the corresponding estimated standard errors associated with fixed effects and random effects regression coefficients; so we have $\hat{Y}_{is}^b = X\hat{\beta}_b + \hat{S}_b$ where b=1~1,000.
2) Generate 1,000 state-level estimates and their standard errors: $\hat{Y}_s^b, \delta_{s,b}^2$; $\delta_{s,b}^2$is the ACS design-based variance for $\hat{Y}_s^b$. It is estimated using the Balanced Repeated Replication (BRR) weighting method that is available in SAS.
3) Generate the final state-level estimates and their standard errors (MSEs): $\bar{Y}_s = \frac{1}{1000}\sum_{b=1}^{1000} \hat{Y}_s^{b'}$ and $var(\bar{Y}_s) = \frac{1}{1000}\sum_{b=1}^{1000}(\hat{Y}_s^{b'} - \bar{Y}_s)^2$; where $\hat{Y}_s^{b'} \sim N(\hat{Y}_s^b, \delta_{s,b}^2)$