The Distribution of Income and Consumption Using Linked Data

Aaron Hong[*]
University of Chicago

Bruce D. Meyer[†]
University of Chicago, NBER, and AEI

Connacher Murphy[‡]
University of Chicago

James X. Sullivan[§]
University of Notre Dame

Derek Wu[**]
University of Virginia

February 5, 2025

Abstract

We examine income and consumption-based measures of well-being and are the first to conduct such analyses using linked expenditure and administrative income data. Our income measures relying on combined, or "blended", survey and administrative data compare much more closely to those of expenditure measures. This is especially true for the very bottom of the distribution, where prior research has revealed concerns about the underreporting of survey income. Blended income deep poverty tends to be very close to consumption-based deep poverty measures, closing almost all of the existing gap in deep poverty rates measured using survey-reported income and consumption.

[*] Email: aahong@uchicago.edu
[†] Email: bdmeyer@uchicago.edu
[‡] Email: connacher.murphy@gmail.com
[§] Email: jsulliv4@nd.edu
[**] Email: derek.wu@virginia.edu

This research was supported by a grant from the U.S. Social Security Administration (SSA) as part of the Retirement and Disability Research Consortium (RDRC). This activity was supported by a contract between the National Academy of Sciences and the Alfred P. Sloan Foundation (No. G-2022-19394) to further the dissemination of the Panel's report, *Creating an Integrated System of Data and Statistics on Household Income, Consumption, and Wealth: Time to Build*. This paper was conducted as a commissioned paper for the report. Any opinions, findings, conclusions, or recommendations expressed in this publication do not necessarily reflect the views of any organization or agencies that provided support for the project. We thank Gerald Auten, David Johnson, four anonymous reviewers, and attendees at the 2024 NTA Conference, 2024 CNSTAT Fall Seminar, and 2025 AEA/ASSA Annual Meetings for helpful comments. The U.S. Census Bureau has reviewed this data product for unauthorized disclosure of confidential information and has approved the disclosure avoidance practices applied to this release, authorization numbers CBDRB-FY23-0184 and CBDRB-FY2023-CES005-016.

1. Introduction

There is a vast literature on the distribution of income, with substantial attention given to low-income families and the role of various government programs in alleviating poverty.[1] However, high and rising rates of underreporting of income in surveys, particularly among those at the bottom of the reported income distribution, raise substantive concerns about bias in studies relying on such measures. Weighted totals from the main Census Bureau household surveys consistently understate national income accounts and government payment totals across various income sources (Meyer, Mok, and Sullivan 2015, Rothbaum 2015). Studies leveraging individual-level linkages between survey and administrative microdata confirm these patterns: 43 percent of Supplemental Nutrition Assistance Program (SNAP) recipients, 39 percent of Unemployment Insurance (UI) recipients, and half of private pension recipients fail to report receipt in the Current Population Survey (CPS), the source of official income and poverty statistics (Bee and Mitchell 2017, Meyer and Mittag 2019, Meyer et al. 2024). Similar patterns of underreporting are observed even in surveys designed to capture more comprehensive income data, such as the Survey of Income and Program Participation (Gathright and Crabb 2014, Meyer, Mittag, and Goerge 2022). This underreporting has important implications for our understanding of poverty, inequality, and the effects of means-tested transfer programs.

Two approaches have emerged to address income underreporting. One approach involves directly linking survey and administrative data to improve income measurement. This method has been especially important in studies of older populations, which have found that incomes of the elderly are often much higher – and the effects of Social Security Administration (SSA) programs more pronounced – than survey data alone suggest (Nicholas and Wiseman 2009, 2010, Bee and Mitchell 2017, Dushi et al. 2024). Research on broader populations similarly finds that the share of individuals below the poverty line and its multiples are much lower after replacing key income components with their administrative equivalents (Meyer and Mittag 2019, Meyer and Wu 2024, Bee et al. 2025). This pattern is particularly true when examining the share below half the poverty line or even lower (Meyer et al. 2021). In addition, the poverty reduction of many social insurance and welfare programs is often understated in survey data (Meyer and Wu 2018).

[1] See, e.g., Ruggles (1990), Citro and Michael (1995), Blank (2002), Hoynes, Page, and Stevens (2006), Ben-Shalom, Moffitt, and Scholz (2012).

A second approach relies on expenditure data, rather than income data, to create measures of consumption. Consumption measures have the advantage of reflecting past savings and debt, the flow of services from owned houses and cars, and the insurance value of government transfers. Since consumption can be smoothed in response to income fluctuations, economists have argued that it provides a better proxy of lifetime utility than annual income (e.g., Poterba 1989, Cutler and Katz 1992). In practice, expenditure data are also less prone to underreporting than income data in surveys (Bee, Meyer, and Sullivan 2015). Numerous studies have documented striking differences between consumption- and income-based measures of inequality and poverty (e.g., Slesnick 1993, Meyer and Sullivan 2012, 2023, Aguiar and Bils 2015, Attanasio and Pistaferri 2016). A consistent finding is that consumption poverty is both lower and has declined more rapidly than income poverty. While a likely explanation for this pattern is the large and growing degree of income underreporting in survey data, no study to date has used individual-level linkages to *directly* quantify the extent to which this underreporting can explain differences in income and consumption poverty.

In this paper, we examine the joint distribution of income- and consumption-based measures of well-being in the United States, applying both of the aforementioned approaches to the same individuals within the same dataset. Specifically, we leverage a unique dataset that links data from the Consumer Expenditure (CE) Survey for 2014-2016 to administrative tax records and program participation data from the Social Security Administration (SSA), Department of Housing and Urban Development (HUD), Department of Veterans Affairs (VA), and state SNAP agencies. This is the first study to undertake such an analysis, with the linked data correcting a substantial portion of income misreporting and having important impacts on the analyses.[2] Indeed, our "blended" income measures, which combine survey and administrative data, compare much more favorably than survey-based aggregates to benchmark totals from national accounts and other sources. Our approach also reflects the recommendations of a recent National Academy of Sciences (2024) report, which emphasized the critical need for comprehensive data integrating well-measured income and consumption.

Our analyses consider measures of both expenditures and consumption, with the former more closely tied to income in an accounting sense (expenditures equal income less saving), while

[2] There is related but quite different international work comparing expenditures derived as a residual from administrative data to survey-reported expenditures (Kreiner et al. 2015, Koijen et al. 2015).

the latter more directly reflecting economic well-being. We begin by examining the full income and expenditure distributions and find that blended income measures align more closely with expenditure distributions than survey-reported income. We also see less evidence of underreporting with the blended income measures, with 21% fewer individuals having expenditures that exceed their income. The changes are especially pronounced at the very bottom of the distribution, where measurement challenges are greatest and government program benefits play a critical role. Notably, the linked data resolve a longstanding puzzle in prior research: individuals reporting the lowest incomes in surveys seemingly report consumption levels many times higher (Meyer and Sullivan 2011, Brewer, Etheridge, and O'Dea 2017). While families in the bottom 5 percent of the survey income distribution report expenditures exceeding average survey income by a factor of 7, this gap narrows to 2.5 when conditioning on the bottom 5 percent of the blended income distribution. Moreover, blended income deep poverty rates closely mirror consumption-based deep poverty rates, closing almost all (97%) of the existing gap measured using survey-reported income and consumption.

2. Data

Our analyses rely on restricted-use data from the interview component of the Consumer Expenditure Survey (CE), which provides the most comprehensive survey data on both household spending and income for a nationally representative sample. These data are linked to administrative tax records from the Internal Revenue Service (IRS) and administrative program records from various government agencies, enabling us to correct measurement errors in survey income components. We focus on data for calendar years 2014-2016, given the availability of both survey and administrative data for those years.

2.1 The Consumer Expenditure Survey

The interview component of the CE is a nationally representative survey of families living in units drawn from the Census Bureau's Master Address File (MAF). The CE is structured as a rotating panel that includes about 7,500 families each quarter. Respondents report detailed spending on a large number of expenditure categories and income from many sources. Families in the survey are interviewed for up to 4 consecutive quarters, with income data typically collected during the first and fourth interviews. Expenditures are reported at the level of the family, or what

the CE refers to as the consumer unit (CU), while incomes are reported at the individual level for those over the age of 14. The CE also contains a diary component that we do not use, as expenditures in the CE Diary Survey compare less favorably to national aggregates than those in the CE Interview Survey (Bee, Meyer, and Sullivan 2015). Surveys are conducted in each of the three months of the quarter, with a reference period of 12 months prior to the interview month for income and 3 months prior for expenditures. For much of the analyses that follow, we compare annual income to four times quarterly expenditures or consumption.[3]

2.2 Administrative Data Sources

From tax records, we obtain adjusted gross income (AGI) reported on IRS 1040 for tax filers. The 1040 extracts also contain information on several other income sources (some of which are components of AGI), such as wages and salaries, interest and dividends, and gross Social Security benefits. Although our 1040 extracts do not contain actual amounts for taxes paid or tax credits received, they cover enough line items (including AGI and filing status) that we can calculate tax liabilities and credits with reasonable accuracy (Meyer et al. 2022). We also rely on IRS Forms W-2 and 1099-R, which are third-party information returns covering wages and retirement distributions, to fill out incomes for individuals who do not file tax returns.

In addition to tax data, we incorporate administrative program records from various federal and state agencies. Social Security (OASDI) benefit amounts are obtained from a combination of universe records from the Master Beneficiary Record (MBR) and Payment History Update System (PHUS) maintained by the SSA (see Wyse et al. 2024c for further details on methods for combining these datasets and extracting benefit amounts). Supplemental Security Income (SSI) amounts are sourced from the SSA's Supplemental Security Record universe file (Wyse et al. 2024a). We also have administrative records on service-connected disability payments (VA DC) to veterans from the VA. OASDI, SSI, and VA DC are all paid out at the individual level.

Housing assistance data are derived from the Public and Indian Housing Information Center (PIC) and Tenant Rental Assistance Certification System (TRACS) files from HUD (Wyse et al. 2024b). These files cover most public and subsidized housing assistance programs under the jurisdiction of HUD, but they miss benefits from programs administered by the Department of

[3] As a robustness check, we also compare annual income to annual expenditures by constructing annual measures of spending summed over four quarters of spending for CUs surveyed for all four quarters.

Agriculture or states and localities. For a subset of results, we also use SNAP data from state agencies, which are available for 22 states during our study period. Housing assistance and SNAP benefits are paid out at the assistance unit level.

3. Linking Data Sources

We link each of these administrative datasets to the CE using Protected Identification Keys (PIKs) generated by the US Census Bureau's Person Identification Validation System (PVS). The PVS is based on a reference file containing Social Security Numbers (SSNs) linked to names, addresses, and dates of birth (Wagner and Layne 2014). PIKs can thus be roughly thought of as scrambled SSNs. While most administrative records have a linkage rate of over 99%, approximately 90% of CUs in our sample contain at least one member linked to a PIK (Appendix Table A2).

Because tax data are only available for calendar years, we restrict our CE sample to survey months where the reference period for income closely aligns with the previous calendar year. Specifically, we include surveys conducted from January through April in survey years 2015 to 2017, corresponding roughly to calendar years 2014 to 2016. The CE includes 77,180 CUs from survey years 2015 to 2017. Restricting to January-April interviews and to first and fourth interviews (during which incomes are reported) yields a sample of 12,898 CUs. Further narrowing the sample to CUs with at least one member with a PIK and an unambiguous state indicator results in a primary analysis sample of approximately 11,200 CUs.[4]

To account for the bias arising from non-random missing PIKs, we divide CU-level survey weights by the predicted probability that at least one CU member has a PIK. Data Appendix A provides further details on the inverse probability weighting procedure that we use to match population totals. Since survey reports of our key concepts (income and expenditures) are incorporated into the weighting estimation, the potential for bias from omitted determinants of these outcomes correlated with being PIKed is small. For un-PIKed individuals (in PIKed CUs) who remain in the sample, we continue to use their survey values in cases where administrative data cannot be linked.

[4] Given small-sample concerns, some observations in the CE are not associated with a particular state. The state indicator is relevant for simulating state tax liabilities and linking administrative SNAP benefits.

3. Defining and Constructing Resource Measures

This section outlines the construction of our primary resource measures. Our main income concept includes both tax liabilities and select in-kind transfers, allowing us to get closer to a measure that approximates the resources available for consumption. We also describe our measure of expenditures and the methods involved in converting this measure to consumption.

3.1 Defining Income

We construct two versions of our post-tax, post-transfer income concept: one based solely on survey data and another using a combination of survey and administrative inputs (which we call blended income). We start with pre-tax money income, which includes: 1) market income sources like earnings, asset income, and retirement income, 2) taxable transfers like OASDI and Unemployment Insurance (UI), and 3) non-taxable transfers like SSI, Temporary Assistance for Needy Families (TANF), and other regular income. To these, we add federal income taxes, state income taxes, and payroll taxes net of the Earned Income Tax Credit (EITC) and Child Tax Credit (CTC). Finally, we include the monetary value of in-kind transfers that support food and housing consumption, specifically rental housing assistance and SNAP.

Calculating the survey income concept is relatively straightforward, as we simply sum the survey-reported amounts corresponding to most income sources. We simulate tax liabilities and credits using the National Bureau of Economic Research's TAXSIM calculator (Feenberg and Coutts 1993), using survey inputs on state, household structure, and incomes. We also impute amounts for survey housing assistance (for which only receipt indicators are reported) by subtracting rent paid from an imputed rental equivalent (see Data Appendix B for details on estimating the rental equivalent). Data Appendix C provides a comprehensive list of the survey variables used for each income source.

Given that constructing the blended income concept is novel and considerably more complex, we devote the next two subsections to describing those methods.

3.2 Constructing Blended Pre-Tax Money Income

We start by describing how we construct a blended version of pre-tax money income. For most CUs that file tax returns, we use Adjusted Gross Income (AGI) reported on IRS Form 1040

as a starting point.[5] AGI provides a relatively comprehensive measure of taxable income, and importantly includes certain income sources for which we do not have individual values in the administrative data (such as alimony, capital gains, and UI). However, AGI has shortcomings. It is net of certain deductions and may miss certain jobs, informal earnings, and other cash sources. As a result, we rely on the survey-reported analog of AGI if it is higher and reflects income that is plausibly missed in the administrative records.

A key challenge is that we cannot perfectly align survey income components with the administrative concept of AGI due to differences in definition (e.g., whether an income source is taxable or non-taxable) and the bundling of survey questions that mix AGI and non-AGI components. We therefore construct a modified version of administrative AGI ("modified AGI") that conceptually matches the union of selected income responses in the survey. This modified AGI serves as a key component of our blended income measure.[6] We continue to use non-imputed survey values of modified AGI when at least one of the following conditions holds:

- Administrative data (from 1040s or W-2s) are missing for the entire CU,
- At least one member of the CU reports self-employment as their primary income source,
- At least one member of the CU reports working in a "high-tip" industry.[7]

This approach stems from the assumption that survey reports of earnings can be both under- and overreported, while administrative earnings data are almost always underreported (see Meyer and Wu 2024). Thus, survey amounts are used to "fill in" holes where we most suspect the administrative values to be incomplete. To validate our methods, we compare our blended amounts for various income sources to publicly available aggregates from national accounts.

For CUs that do not file tax returns, administrative AGI is unavailable. In these cases, we begin with the survey measure of modified AGI and make component-by-component substitutions using wages/salaries from W-2s and retirement distributions from 1099-Rs to construct a blended

[5] In other work (e.g., Meyer and Wu 2024) using the CPS ASEC, the authors start with the survey income concept and replace administrative or blended values component-by-component. We use a different approach here – starting with an aggregate administrative concept like AGI – for two reasons. First, several administrative datasets available for the CPS are not available for the CE, such as the SSA's Detailed Earnings Record which contains self-employment amounts. These amounts, however, are contained in AGI. Second, a number of survey income concepts are bundled in the CE, making it hard to do component-by-component adjustments when one of those bundled components is missing in the administrative data.

[6] Using administrative data, modified AGI is simply AGI plus non-taxable interest (from IRS Form 1040), deferred compensation for wages/salaries (from IRS Form W-2), and veterans' benefits (from administrative VA data) minus taxable OASDI. Data Appendices D and E provide more details on the construction of modified AGI.

[7] We define high-tip industries as sales, retail, private household service, construction, mechanics, and farming.

measure of modified AGI. After constructing a blended measure of modified AGI for all CUs, we bring in additional money income sources that have not yet been explicitly accounted for. These include administrative amounts for OASDI and SSI and survey values for TANF. Data Appendix D provides further details on constructing blended amounts for each income source.

3.3 Constructing Blended Taxes and In-Kind Transfers

To calculate blended measures of tax liabilities, we use TAXSIM and apply inputs from a combination of IRS tax records and other sources. AGI and tax unit structure (e.g., filing status, number of dependents) are derived from the 1040s, and we estimate the number of qualifying dependents for the EITC and CTC by linking birth dates from the SSA's Numident file. We calculate payroll taxes on 1) the higher of W-2 and survey wages (since Social Security taxes are capped at the individual level and payroll taxes are collected for non-filers) and 2) survey-reported self-employment income (as we lack component-level administrative data on self-employment income). We also calculate taxes accrued for those who do not appear in 1040s (in the event they are unlinked, filed late, or had taxes withheld), using their survey-reported family structure and on incomes from either IRS information returns or the survey.

We subsequently bring in amounts corresponding to housing assistance and SNAP. For housing assistance, given that the administrative amounts cover only HUD-administered programs, we treat survey-reported recipients who do not appear in the administrative data as true recipients. Given that HUD programs are understated in survey data (Meyer and Wu 2018, Meyer and Mittag 2019), we may continue to understate housing assistance overall in light of the likely false negatives associated with non-HUD programs that we cannot correct. For SNAP, we continue to use survey-reported SNAP for our main analyses, although we bring in administrative data for 22 states as part of our robustness checks.

3.4 Constructing Expenditures and Consumption

Our measure of expenditures includes all spending reported in the CE Interview survey. To convert reported expenditures into a measure of consumption, we apply several adjustments (see Data Appendix A for more details). First, we convert vehicle spending to a service flow equivalent, which is a function of a depreciation rate and the current market value of the vehicle. Second, to convert housing expenditures to housing consumption for homeowners, we substitute

the reported rental equivalent of the home for the sum of mortgage interest payments, property tax payments, spending on insurance, and maintenance and repairs. Third, we impute a rental equivalent for those who report living in government or subsidized housing using reported information on their living unit, including the number of rooms, bedrooms and bathrooms, and the presence of appliances.[8] Finally, we exclude spending that is better interpreted as an investment (spending on education, health care, and outlays for retirement) or is consumed by those outside the CU (charitable contributions and gifts to non-family members). We also exclude out-of-pocket medical expenses because high out-of-pocket expenses may reflect substantial need or lack of good insurance, rather than greater well-being.

4. Comparisons of Weighted Totals to Public Aggregates

Before presenting our main results, we compare the weighted totals of blended and survey income components to benchmarks from publicly available aggregates, such as the National Income and Product Accounts and other sources. This exercise highlights the extent of underreporting in aggregate survey income and serves as a validation exercise for our blended income methodology.

Figure 1 shows three-year average reporting rates for key income components, revealing that all survey-based income measures suffer from underreporting.[9] For instance, survey-based modified AGI captures only 73% of its aggregate counterpart. While the reporting rate for the single largest component of survey-based modified AGI – wage and salary income – is fairly high (92%), the reporting rates vary widely across other components. For example, the rate is high for survey OASDI (88%) but low for other income sources such as retirement and pensions (27%), other regular income (31%), and SSI (49%).[10]

The reporting rates for blended income measures are not only substantially higher but also close to one in many cases. The reporting rate for blended modified AGI is 97%, and the rate for blended wage and salary income is 103%. However, some components of blended income continue to have low reporting rates. For example, the rate for our blended measure of other regular income

[8] Since housing assistance is underreported in the CE, this will be a reason why our measure of consumption will be slightly understated.
[9] Yearly reporting rates for other components of income can be found in Appendix Tables A4, A5a, A5b, and A5c.
[10] These low reporting rates for income components are not unique to the CE. These components also have low reporting rates in other major surveys including the CPS, the ACS, and the SIPP (Meyer, Mok, and Sullivan 2015).

(where we must still rely on survey data for many pieces) is only 56% of the administrative aggregate. On net, Figure 1 suggests that our blended income measure addresses much of the underreporting evident in survey income and thus provides a more accurate picture of economic well-being.[11] However, our blended income measure still misses administrative records for a number of income sources that are likely to be underreported in surveys, such as workers' compensation, child support, and TANF.

5. Distributions of Expenditures and Income

To further explore the nature and implications of underreported survey income, we examine the distributions of our income measures and compare them to those of expenditures.[12] This allows us to assess the extent to which blended income narrows the gap between income and expenditures, particularly for families with few resources. We start by comparing the univariate distributions of these measures before exploring their joint distribution and the individual-level discrepancies between income and expenditures. To account for differences in family size and composition, we scale all measures using the Supplemental Poverty Measure (SPM) three-parameter equivalence scale (see Data Appendix F). Dollar figures are expressed in real 2016 dollars, adjusted using the Personal Consumption Expenditures Price Index (PCEPI).

5.1 Univariate Distribution of Resource Measures

Figure 2 plots the bottom fifty percentiles for three resource measures – blended income, survey income, and total expenditures – pooled across the three years in our sample (see also rows 1, 4, and 7 of Table 1).[13] Given the aggregate comparisons in the prior section, it is not surprising that blended income exceeds survey income at many points of the distribution. However, Figure 3 shows that this gap persists throughout the *entire* bottom half, with the level difference growing as one moves up the distribution. Consistent with earlier research using CE data (Meyer and Sullivan 2011), expenditures substantially exceed survey income at very low percentiles. For example, at the 3rd percentile, total expenditures are 140% higher than survey income while only

[11] Appendix Figure A1 also shows that the reporting rates for blended tax components are quite high – all but one has a rate above 90%.

[12] For this section, we omit housing assistance from income to enable an apples-to-apples comparison with expenditures, which do not include subsidized rent.

[13] See Figure A2 for the top half of the distribution.

6% higher than blended income. Furthermore, while expenditures exceed survey income for the bottom ten percentiles, they only exceed blended income for the bottom three percentiles.

5.2 Joint Distribution of Resource Measures

Differences between income and expenditures at specific percentiles can arise because families at a given point in one distribution are not necessarily at the same point in another distribution. To compare income and expenditures for the same families, we examine the joint distribution of these resource measures. Table 1 reports the means of resource measures for families below specified percentiles of each resource measure.

Focusing on families in the bottom 5% of the survey income distribution, we find that mean expenditures (row 9) exceed mean survey income (row 2) by a factor of 7 – a difference consistent with evidence provided in previous studies (Meyer and Sullivan 2011, Brewer, Etheridge, and O'Dea 2017). That expenditures exceed income is unsurprising, because we are conditioning on low survey income while not restricting to low expenditures. Reversing the exercise – conditioning on low expenditures – reveals a much smaller gap. For families in the bottom 5% of the expenditure distribution, mean survey income (row 3) exceeds mean expenditures (row 8) by a factor of 1.75.

Blended income substantially reduces these discrepancies. For those in the bottom 5% of the blended income distribution, mean expenditures (row 10) exceed mean blended income (row 5) by only a factor of 2.5. While nearly two-thirds of the discrepancy is resolved by using blended income, unreported family transfers and savings likely remain as important possible explanations. This pattern suggests that income underreporting can explain much of the observed disparity between expenditures and survey income in the left tail.

Given the strong relationship between income and expenditures, especially for families with little borrowing and saving, we would generally expect expenditures to rise with income. However, Table 1 shows that expenditures do not increase monotonically with survey income at low levels. In fact, average spending below the 5th percentile of survey income is 6.8% greater than average spending below the 10th percentile of survey income (row 9). This non-monotonicity is less evident when examining spending by *blended* income (row 10). Below the 10th percentile of blended income, expenditures are largely flat. Those below the 3rd percentile of blended income spend 1% more than those below the 5th percentile, but expenditures for families below the 5th percentile of blended are slightly lower than for those below the 10th percentile. This pattern

suggests that blended income, which incorporates both survey and administrative data, is subject to less measurement error and provides a better indicator of economic well-being at the bottom of the distribution.

A similar relationship is evident in Figure 3a, which plots median expenditures in the bottom half of the survey and blended income distributions. Specifically, we show median expenditures by bins of income, where each bin includes approximately 3 percentiles. For example, the leftmost point on the figure represents median expenditures of $25,900 for families in the bottom 3 percentiles of the survey income distribution. Figure 3a again highlights the non-monotonicity of expenditures at the bottom of the survey income distribution, notably between the second and third data points (6th and 9th percentiles) and between the fourth and fifth data points (12th and 15th percentiles). In contrast, the trend in median expenditures by blended income is monotonic until a substantially higher percentile (45th).

5.3 Individual-Level Differences in Resource Measures

We also analyze within-family differences between income and expenditures to assess how these discrepancies change when using blended income instead of survey income. This exercise further illustrates how improving the measurement of income reduces the mismatch between reported income and expenditures, particularly when the latter exceeds the former. Figure 3b plots the distribution of differences between 1) expenditures and survey income and 2) expenditures and blended income for all individuals in our sample. Reported expenditures exceed survey incomes for 37% of individuals, with this share falling by 21% (to 29%) when comparing expenditures to blended income. This is consistent with a leftward shift in the distribution of expenditures minus income when survey income is replaced with blended income.

We replicate this analysis across various subsamples and robustness checks (Appendix Figure A3). Across all groups, we consistently observe fewer individuals whose expenditures exceed income after switching from survey to blended income. Specifically, we observe a 33.2% drop for those 65 and over, a 35.4% drop for families reporting less than $50,000 in survey income, a 22.7% drop for residents of states with available administrative SNAP data, and a 27.5% drop for those interviewed in the fourth wave of the CE with at least one prior interview. In each subsample, using blended income continues to lead to a leftward shift in the distribution of expenditures minus income.

Given that expenditures should equal the sum of after-tax income and saving or dissaving, any differences between expenditures and income should be attributed to either income misreporting or saving behavior.[14] The considerable differences between income and expenditures that remain at the lower end of the income distribution (where there is relatively little saving and dissaving) point to the continued persistence of measurement error – potentially due to blended income still missing incomes for which we do not have administrative sources.

6. Income and Consumption Poverty Measures

Our analyses thus far have compared the distributions of income and expenditures, showing that blended income exceeds survey income at every percentile and better matches expenditures at the bottom of the distribution. In this section, we focus on applications to deep poverty and poverty, which summarize the resource distribution at particular thresholds. Specifically, we calculate the share of individuals in CUs below the SPM poverty line and half the poverty line using survey income, blended income, and consumption measures.[15] All resource measures are adjusted using the SPM three-parameter equivalence scale and normalized to a family of two adults and two children.[16] While we previously examined expenditures, we now focus on consumption because it conceptually aligns more closely with economic well-being. Whereas expenditures can be impacted by large irregular purchases like a down payment on a house or a vehicle purchase, consumption reflects the smoothed flow of services obtained from these durable goods.

Deep Poverty

We begin by examining how the deep poverty rate, defined as the share of those with resources below half the poverty line (Figure 4a). Families near this threshold typically have limited saving or borrowing, meaning measurement error likely accounts for much of the discrepancies between income and consumption. Using survey income, 5.3% of all individuals are

[14] Alternatively, there could be misreporting of expenditures, but to explain why expenditures exceed income at the bottom, expenditures would have to be over-reported, and previous studies have found little evidence of overreporting (Bee, Meyer, and Sullivan 2015, Brewer, Etheridge, and O'Dea 2017).

[15] To establish our poverty line, we take the weighted average of the SPM threshold, which varies by housing tenure, for reference year 2016 (Fox 2017). To extend the poverty line back to 2014 and 2015, we deflate this value using the PCEPI. We then calculate the overall poverty rate by averaging across the three years of our data.

[16] At half the poverty line, the value of housing benefits is not included in our measure of survey income. However, at the poverty line, we do include the value of housing benefits.

classified as deep poor. This rate falls to 1.9% with blended income, while the deep poverty rate using reported consumption is 1.8%. In other words, 97% of the gap between survey-reported rates of consumption and income deep poverty can be attributed to income underreporting. Moreover, the similarity of the deep poverty rates for blended income and consumption (differing by only 0.1 percentage points) suggests that using consumption to identify the most disadvantaged may yield similar results as those of an income measure that corrects for misreporting.

We also estimate deep poverty for selected (overlapping) demographic subgroups: individuals in non-elderly single-parent families, children (under age 18), and the elderly (age 65 and older). For all subgroups, the underreporting of survey income continues to explain a large fraction of the gap between survey-reported income and consumption deep poverty. Using blended income closes this gap by more than 100% for children, 87% for the elderly, and 76% for single-parent families. While differences between blended income and consumption deep poverty still persist for the elderly and single parents, they are not unexpected. The differences for the elderly can be explained by the higher likelihood of increased dissaving and asset ownership for this group. For single parents, our blended income measure does not include administrative data for a number of programs that are heavily underreported in the CE but important for low-income single parents (e.g., TANF, WIC, and child support).

Poverty

Moving to higher thresholds, Figure 4b examines the share of individuals in CUs with resources below the poverty line. For the full sample, 13% of individuals have survey incomes below the poverty line, compared to 7.3% for blended income – a 44% reduction. In contrast, the consumption poverty rate is 17.9%, exceeding both survey and blended income poverty rates. We should expect the difference between consumption and income measures to grow further up the distribution, as families with more resources are more likely to save. Some of this difference could also be due to greater underreporting of expenditures higher in the distribution.

For all three demographic subgroups, survey income poverty rates exceed those of blended income, with the gap most pronounced for the elderly. For this group, the blended income poverty rate (5.5%) is 53% lower than that of survey income poverty (11.6%), largely due to the underreporting of retirement and pension income in the survey (Figure 1). Consumption poverty

rates are also substantially higher than blended income poverty across all subgroups, although the magnitude of this difference varies by groups.

In the results discussed so far, our blended income measure relies on survey-reported SNAP benefits due to the limited availability of administrative SNAP records. However, the reduction in income poverty is 49% after incorporating administrative SNAP records (compared to 44% without administrative SNAP), highlighting the impact of correcting for SNAP underreporting (Appendix Table A7). At higher poverty thresholds (150% and 200% of the poverty line), blended income still results in lower poverty rates than survey income, although the relative differences diminish (Appendix Table A6). Consumption poverty consistently exceeds income poverty at higher thresholds, reflecting a greater propensity to save further up the distribution. The underreporting of expenditures may also matter more higher up in the distribution.

7. Discussion and Conclusions

This paper constructs the first integrated dataset that combines accurate income and consumption measures of family resources, using survey values of income and expenditures linked to comprehensive administrative tax and program data. By blending survey and administrative data, we correct for the misreporting of key income sources – such as earnings, OASDI, retirement pensions, and SSI – while retaining demographic detail, family structure, and information on non-administrative income sources.

A key finding is that the distributions of income and expenditures align much more closely when using blended rather than survey income. At the very bottom, using blended income reduces the deep poverty rate to 1.9%, closely matching the consumption-based deep poverty rate of 1.8%. In contrast, the survey-reported income deep poverty rate is markedly higher at 5.3%. These findings have two key implications. First, survey-reported income overstates poverty, particularly among those with severe material deprivation. Second, consumption provides a meaningful proxy for income corrected for measurement error, especially at the bottom of the distribution.

Our findings highlight the critical value of a data infrastructure that links survey and administrative data, enabling more precise measurement of economic well-being by reducing misreporting (see NASEM 2024). Such data integration can pave the way for innovative methodologies that inform evidence-based policymaking and advance understanding of inequality

and poverty dynamics. Future research should also prioritize incorporating measures of wealth and savings to better reconcile differences between income and consumption and provide a more complete picture of household economic dynamics.

References

Aguiar, Mark and Mark Bils. 2015. "Has Consumption Inequality Mirrored Income Inequality?" *American Economic Review*, 105(9): 2725-2756.

Attanasio, Orazio P. and Luigi Pistaferri. 2016. "Consumption Inequality." *Journal of Economic Perspectives*, 30(2): 3-28.

Bee, C. Adam, Bruce Meyer, and James Sullivan. 2015. "The Validity of Consumption Data: Are the Consumer Expenditure Interview and Diary Surveys Informative?" In Carroll, C., T. Crossley, and J. Sabelhaus (eds.), *Improving the Measurement of Consumer Expenditures*, 204-240. Chicago, IL: University of Chicago Press.

Bee, Adam, John Creamer, Joshua Mitchell, Nikolas Mittag, Jonathan Rothbaum, Carl Sanders, Lawrence Schmidt, and Matthew Unrath. 2025. "National Experimental Well-Being Statistics Technical Note on Methodological Changes for Version 2." SESHD Working Paper No. FY2025-01. Washington, DC: U.S. Census Bureau.

Bee, C. Adam and Joshua Mitchell. 2017. "Do Older Americans Have More Income Than We Think?" SEHSD Working Paper No. 2017-39. Washington, DC: U.S. Census Bureau.

Ben-Shalom, Yonatan, Robert Moffitt, and John Karl Scholz. 2012. "An Assessment of the Effectiveness of Anti-Poverty Programs in the United States." In P. N. Jefferson (ed.), *The Oxford Handbook of the Economics of Poverty*, 709-749. Oxford, UK: Oxford

Blank, Rebecca M. 2002. "Evaluating Welfare Reform in the United States." *Journal of Economic Literature*, 40(4): 1105-1166.

Brewer, Mike, Ben Etheridge and Cormac O'Dea. 2017. "Why are Households that Report the Lowest Incomes So Well-off?" *The Economic Journal*, 127(605): F24-F49.

Citro, Constance F. and Robert T. Michael, eds. 1995. *Measuring Poverty: A New Approach.* Washington, DC: National Academy Press.

Corinth, Kevin, Bruce D. Meyer, and Derek Wu. 2022. "The Change in Poverty from 1995 to 2016 among Single-Parent Families." *AEA Papers and Proceedings*, 112: 345-350.

Cutler, David M. and Lawrence F. Katz. 1992. "Rising Inequality? Changes in the Distribution of Income and Consumption in the 1980's." *American Economic Review*, 82(2): 546–551.

Dushi, Irena, Brad Trenkamp, Charles Adam Bee, and Joshua W. Mitchell. 2024. "Measuring Income of the Aged in Household Surveys: Evidence from Linked Administrative Records." *Journal of Pension Economics and Finance*, 24(1): 95-122.

Feenberg, Daniel, and Elisabeth Coutts. 1993. "An Introduction to the TAXSIM Model." *Journal of Policy Analysis and Management*, 12(1): 189-194.

Fox, Liana. 2017. "The Supplemental Poverty Measure: 2016". Report No. P60-261. Washington, DC: U.S. Census Bureau.

Gathright, Graton and Tyler A. Crabb. 2014. "Reporting of SSA Program Participation in the SIPP." Working Paper.

Han, Jeehoon, Bruce D, Meyer and James X. Sullivan. 2021. "The Consumption, Income, and Well-Being of Single Mother Headed Families 25 Years After Welfare Reform." *National Tax Journal*, 74(3): 791-824.

Hoynes, Hilary W., Marianne E. Page, and Ann Huff Stevens. 2006. "Poverty in America: Trends and Explanations." *Journal of Economic Perspectives*, 20(1): 47-68.

Koijen, Ralph, Stijn Van Nieuwerburgh, and Roine Vestman. 2015. "Judging the Quality of Survey Data by Comparison with 'Truth' as Measured by Administrative Records: Evidence From Sweden." In Carroll, C., T. Crossley, and J. Sabelhaus (eds.), *Improving the Measurement of Consumer Expenditures*, 308-346. Chicago, IL: University of Chicago Press.

Kreiner, Claus Thustrup, David Dreyer Lassen, Søren Leth-Petersen. 2015. "Measuring the Accuracy of Survey Responses Using Administrative Register Data: Evidence from Denmark." In Carroll, C., T. Crossley, and J. Sabelhaus (eds.), *Improving the Measurement of Consumer Expenditures*, 289-307. Chicago, IL: University of Chicago Press.

Meyer, Bruce D. and Nikolas Mittag. 2019. "Using Linked Survey and Administrative Data to Better Measure Income: Implications for Poverty, Program Effectiveness and Holes in the Safety Net." *American Economic Journal: Applied Economics*, 11(2): 176-204.

Meyer, Bruce D., Nikolas Mittag and Robert Goerge. 2022. "Errors in Survey Reporting and Imputation and their Effects on Estimates of Food Stamp Program Participation." *Journal of Human Resources*, 57(5): 1605-1643.

Meyer, Bruce D., Wallace K. C. Mok, and James X. Sullivan. 2015. "Household Surveys in Crisis." *Journal of Economic Perspectives*, 29(4): 199–226.

Meyer, Bruce D. and James X. Sullivan. 2003. "Measuring the Well-Being of the Poor Using Income and Consumption." *Journal of Human Resources*, 38(Supplement): 1180-1220.

———. 2011. "Further Evidence on Measuring the Well-Being of the Poor Using Income and Consumption." *Canadian Journal of Economics*, 44(1): 52-87.

———. 2012. "Identifying the Disadvantaged: Official Poverty, Consumption Poverty, and the New Supplemental Poverty Measure." *Journal of Economic Perspectives*, 26(3): 111-136.

———. 2023. "Consumption and Income Inequality in the U.S. Since the 1960s." *Journal of Political Economy*, 131(2): 247-284.

Meyer, Bruce D., Connacher Murphy, and James X. Sullivan. 2022. "Changes in the Distribution of Economic Well-Being During the COVID-19 Pandemic: Evidence from Nationally Representative Consumption Data." NBER Working Paper No. 29878.

Meyer, Bruce D. and Derek Wu. 2018. "The Poverty Reduction of Social Security and Means-Tested Transfers." *Industrial and Labor Relations Review*, 71(5): 1106-1153.

———. 2024. "Poverty in the United States." Working Paper.

Meyer, Bruce, Derek Wu, Victoria R. Mooers, and Carla Medalia. 2021. "The Use and Misuse of Income Data and Extreme Poverty in the United States." *Journal of Labor Economics*, 39(S1): S5-S58.

Meyer, Bruce D., Derek Wu, Grace Finley, Patrick Langetieg, Carla Medalia, Mark Payne and Alan Plumley. 2022. "The Accuracy of Tax Imputations: Estimating Tax Liabilities and Credits Using Linked Survey and Administrative Data." In Chetty, R., J. Friedman, J. Gornick, B. Johnson, and A. Kennickell (eds.), *Measuring Distribution and Mobility of Income and Wealth*, 459-498. NBER Book Series.

Meyer, Bruce D., Derek Wu, Matthew Stadnicki, and Patrick Langetieg. 2024. "The Reporting of Unemployment Insurance and Unemployment in Survey and Administrative Sources." Working Paper.

National Academies of Science, Engineering and Medicine (NASEM). 2024. *Creating an Integrated System of Data and Statistics on Household Income, Consumption, and Wealth: Time to Build.* Washington, DC: The National Academies Press.

Nicholas, Joyce and Michael Wiseman. 2009. "Elderly Poverty and Supplemental Security Income." *Social Security Bulletin*, 69(1): 45-73.

———. 2010. "Elderly Poverty and Supplemental Security Income, 2002-2005." *Social Security Bulletin*, 70(2): 1-29.

Poterba, James M. 1989. "Lifetime Incidence and the Distributional Burden of Excise Taxes." *American Economic Review*, 79(2): 325–330.

Rothbaum, Jonathan L. 2015. "Comparing Income Aggregates: How do the CPS and ACS Match the National Income and Product Accounts, 2007–2012." SESHD Working Paper No. 2015-01. Washington, DC: U.S. Census Bureau.

Ruggles, Patricia. 1990. *Drawing the Line: Alternative Poverty Measures and their Implications for Public Policy*. Washington, DC: Urban Institute Press.

Slesnick, Daniel T. 1993. "Gaining Ground: Poverty in the Postwar United States." *Journal of Political Economy*, 101(1): 1-38.

Wagner, Deborah, and Mary Lane. 2014. "The Person Identification Validation System (PVS): Applying the Center for Administrative Records Research and Applications' (CARRA) Record Linkage Software." CARRA Working Paper No. 2014-01. Washington, DC: U.S. Census Bureau.

Wyse, Angela, Gillian Meyer, Derek Wu, and Bruce D. Meyer. 2024a. "Measuring Income from SSI Benefits with Administrative Universe Data." CID Technical Note Series 2024-01.

_____. 2024b. "Measuring Longitudinal HUD Benefit Receipt with Administrative Data." CID Technical Note Series 2024-03.

Wyse, Angela, Connacher Murphy, Derek Wu, and Bruce D. Meyer. 2024c. "Measuring Income from OASDI Benefits with Administrative Universe Data." CID Technical Note Series 2024-02.

Tables and Figures

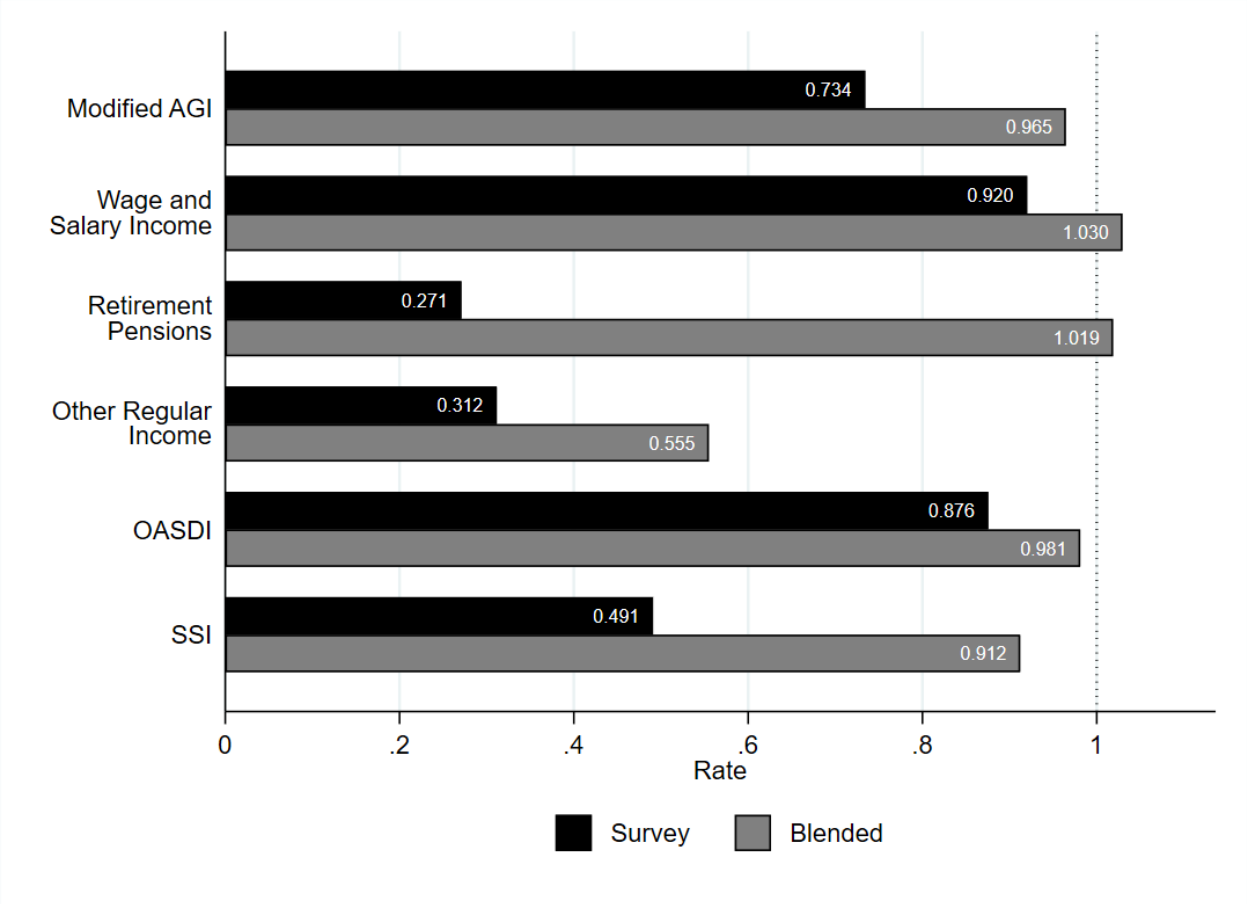
Table 1. Percentiles and Conditional Means of Resource Measures

		Percentiles								
		3rd	5th	10th	20th	30th	40th	50th	75th	90th
		Survey Income								
(1)	Percentile of survey income	6,448	12,610	22,250	32,150	40,660	49,720	59,620	93,550	143,600
(2)	Mean for those below given percentile of survey income	2,501	5,287	11,610	19,520	25,260	30,380	35,270	48,580	59,800
(3)	Mean for those below given percentile of expenditures	21,940	24,110	27,200	33,140	36,870	41,050	45,200	57,120	66,080
		Blended Income								
(4)	Percentile of blended income	14,860	20,790	28,680	40,370	50,940	63,420	75,980	119,200	188,200
(5)	Mean for those below given percentile of blended income	8,282	12,390	18,830	26,710	33,070	39,120	45,240	62,000	76,620
(6)	Mean for those below given percentile of expenditures	33,020	37,170	39,880	46,600	50,700	55,580	60,210	74,540	86,420
		Expenditures								
(7)	Percentile of expenditures	15,710	17,820	22,400	29,420	35,730	42,360	49,430	75,990	119,800
(8)	Mean for those below given percentile of expenditures	11,840	13,780	17,100	21,640	25,300	28,790	32,240	42,070	51,050
(9)	Mean for those below given percentile of survey income	34,270	36,950	34,590	35,560	37,240	40,140	43,080	50,230	57,530
(10)	Mean for those below given percentile of blended income	31,340	31,060	31,440	33,480	36,650	39,660	42,630	50,430	57,520

Source: 2014-16 Consumer Expenditure Survey linked to various administrative records

Notes: This table reports reports the means of resource measures for families below specified percentiles of each resource measure. All values are expressed in 2016 dollars and adjusted for differences in family size using the three-parameter SPM equivalence scale and normalized to a two-adult and two-children family. Income is after tax and includes food stamps. Both income and expenditures exclude housing assistance. Approved for release by the Census Bureau's Disclosure Review Board, authorization numbers CBDRB-FY23-0184 and CBDRB-FY2023-CES005-016.

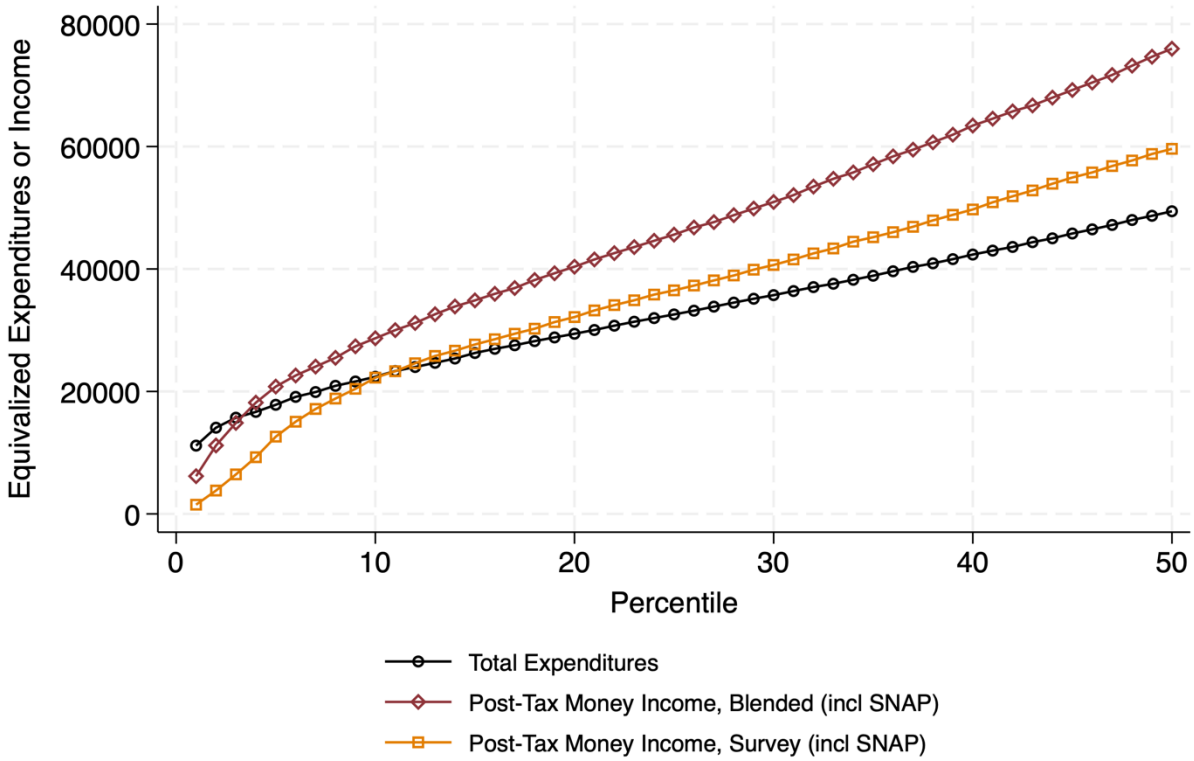
Figure 1. Average Reporting Rates of Blended vs. Survey Income Categories (2014-2016)



Source: 2014-16 Consumer Expenditure Survey, publicly available aggregates from NIPA and other sources

Notes: This figure presents the average of yearly reporting rates for key income components. We compare the weighted totals of survey and blended income components to publicly available aggregates. A description of the weighting methodology can be found in Data Appendix A and the sources used to construct publicly available aggregates are listed in Appendix Table A3. The underlying numbers can be found in Appendix Tables A4, A5a, and A5b. Modified AGI is defined as the sum of wages and salary income, self-employment income, retirement pensions, interest and dividends, rental income and royalties, other regular income, and other non-rental income. While the survey definition of other regular income consists of income components such as VA benefits, worker's compensation, UI, child support, and alimony, we only have the corresponding administrative sources for VA benefits. Hence, our blended measure is relatively understated for other regular income. Approved for release by the Census Bureau's Disclosure Review Board, authorization numbers CBDRB-FY23-0184 and CBDRB-FY2023-CES005-016.

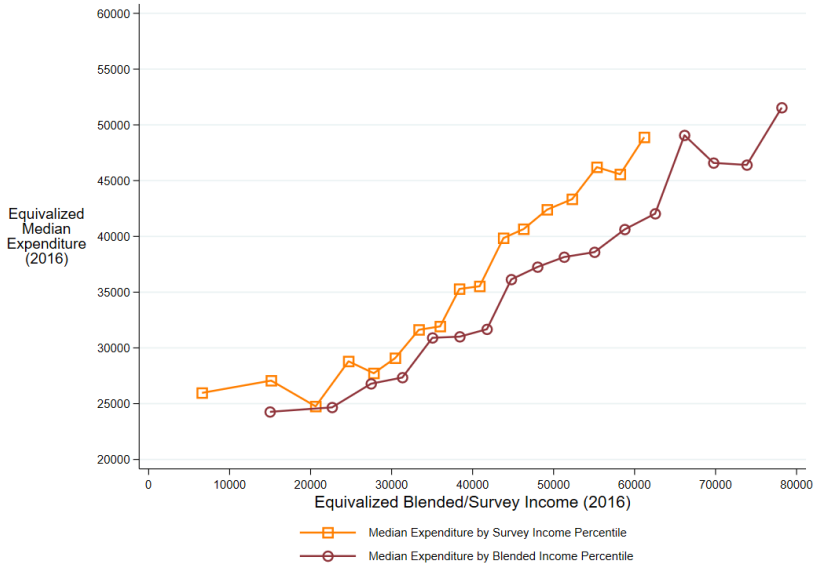
Figure 2. Distribution of Resource Measures, Bottom 50 Percentiles



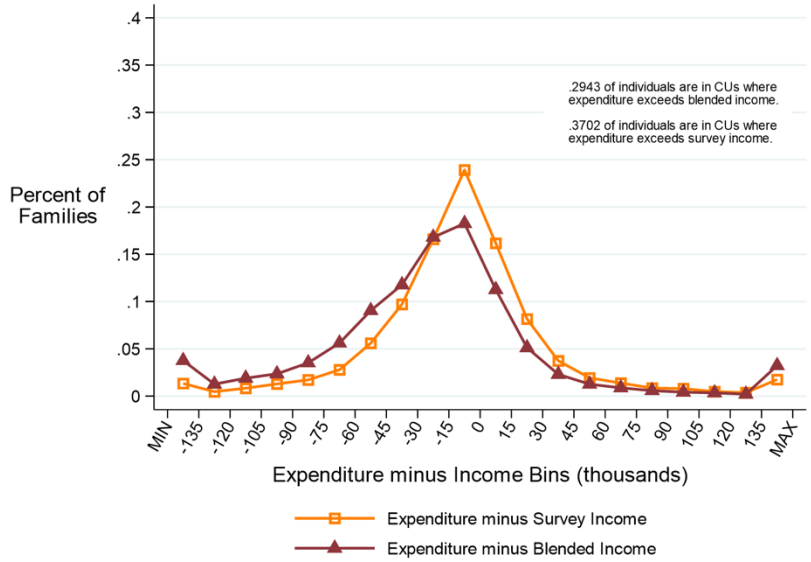
Source: 2014-16 Consumer Expenditure Survey linked to various administrative records

Notes: This figure shows the bottom fifty percentiles for three resource measures – blended income, survey income, and total expenditures – pooled across the three years in our sample. All values are expressed in 2016 dollars and adjusted for differences in family size using the three-parameter SPM equivalence scale and normalized to a two-adult and two-child family. Income is after tax and includes survey-reported SNAP. Both income and expenditures exclude housing assistance. Approved for release by the Census Bureau's Disclosure Review Board, authorization numbers CBDRB-FY23-0184 and CBDRB-FY2023-CES005-016.

Figure 3. Measures of the Joint Distribution of Income and Expenditure



(a) Median Expenditure by Income Quantile

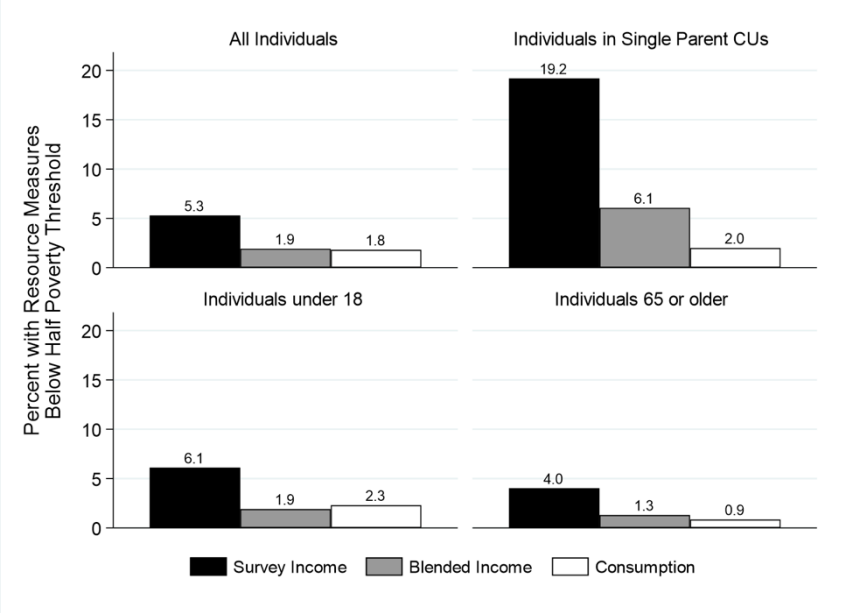


(b) Distribution of Expenditure minus Income

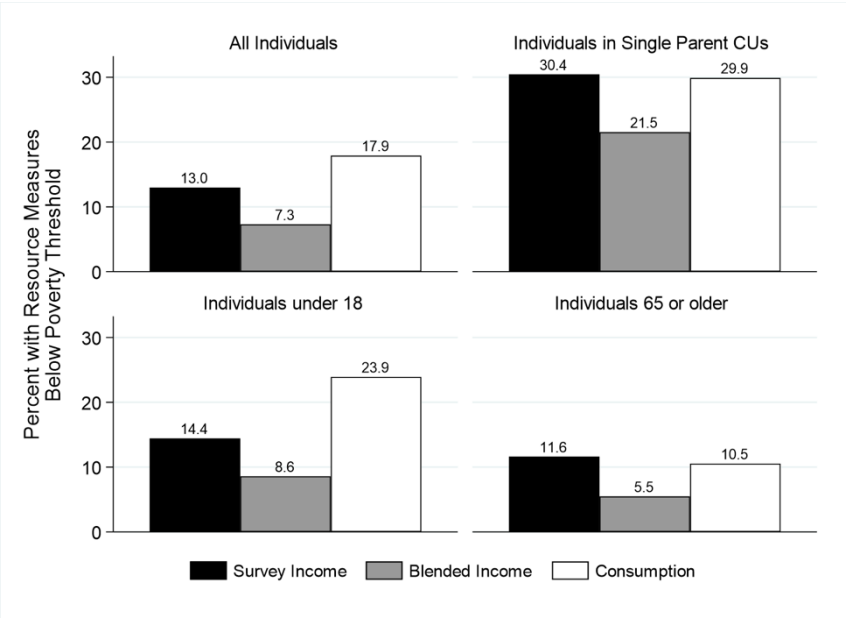
Source: 2014-16 Consumer Expenditure Survey linked to various administrative records

Notes: Panel (a) plots median expenditures in the bottom half of the survey and blended income distributions, after dividing the sample into 33 equally populated bins and displaying the bottom 17 bins. Panel (b) plots the distribution of differences between 1) expenditures and survey income and 2) expenditures and blended income, after partitioning the data into $15,000 bins and positioning markers at the midpoints of these bins. All values are expressed in 2016 dollars and adjusted for differences in family size using the three-parameter SPM equivalence scale and normalized to a two-adult and two-child family. Income is after tax and includes survey-reported SNAP. Both income and expenditures exclude housing assistance. Approved for release by the Census Bureau's Disclosure Review Board, authorization numbers CBDRB-FY23-0184 and CBDRB-FY2023-CES005-016.

Figure 4. Share of Individuals with Resources Below Half Poverty Line and Poverty Line



(a) Below Half Poverty Line (Deep Poverty)



(b) Below Poverty Line (Regular Poverty)

Source: 2014-16 Consumer Expenditure Survey linked to various administrative records

Notes: These figures show the share of individuals in consumer units with resources below 50% of the poverty line (Panel a) and 100% of the poverty line (Panel b). We show estimates for all individuals, along with those in single parent CUs, those under the age of 18, and those aged 65 and over. Both income and consumption measures include housing assistance. The survey and blended income concepts are post-tax and include housing assistance and survey-reported SNAP. All resource measures have been adjusted for inflation, equivalized using the SPM three-parameter equivalence scale, and normalized to a family consisting of two adults and two children. Approved for release by the Census Bureau's Disclosure Review Board, authorization numbers CBDRB-FY23-0184 and CBDRB-FY2023-CES005-016.

Appendix Tables and Figures

Table A1. Sample Sizes

Sample Subset Type (1)	Sample Subset (2)	Unweighted Number of CUs (3)	Unweighted Number of Individuals (4)
	Panel A. 1st and 4th Interview, Full Sample		
Year	Reference Year 2014	3,500	8,700
Year	Reference Year 2015	3,900	9,600
Year	Reference Year 2016	3,800	9,200
Family Type	Single Adult w/ no Kids (or Just Kids)	2,000	2,000
Family Type	Single Adult w/ Kids	550	1,450
Family Type	Multiple Adults w/ Kids	2,800	12,000
Family Type	Multiple Adults w/ no Kids	2,600	6,000
Family Type	Families containing 65+ Adults	3,300	6,400
Age	Age: Less than 18		6,700
Age	Age: 18-54		13,000
Age	Age: 55-64		3,600
Age	Age: 65 or older		4,500
	Panel B. 1st and 4th Interview, States with Administrative SNAP Data		
Year	Reference Year 2014	1,500	3,700
Year	Reference Year 2015	1,600	3,900
Year	Reference Year 2016	1,500	3,800
Family Type	Single Adult w/ no Kids (or Just Kids)	800	800
Family Type	Single Adult w/ Kids	200	550
Family Type	Multiple Adults w/ Kids	1,100	4,700
Family Type	Multiple Adults w/ no Kids	1,100	2,500
Family Type	Families containing 65+ Adults	1,500	2,900
Age	Age: Less than 18		2,700
Age	Age: 18-54		5,200
Age	Age: 55-64		1,500
Age	Age: 65 or older		2,000

Source: 2014-16 Consumer Expenditure Survey linked to various administrative records

Notes: This table presents the unweighted sample sizes for CUs that include at least one member with a PIK and have an unambiguous state indicator and are interviewed in the 1st and 4th wave between January and April of calendar years 2015-2017. Counts of observations at the CU level and the member level are included. Values are rounded in accordance with Census disclosure rules. All results were approved for release by the Census Bureau's Disclosure Review Board, authorization numbers CBDRB-FY23-0184 and CBDRB-FY2023-CES005-016.

Table A2. PIK Rates

Reference Year	Consumer Unit-Level			Individual-Level		
	Number of CUs with at least 1 PIKed Member	Total Number of CUs	PIK Rate	Number of Individuals in CUs with at least 1 PIKed Member	Total Number of Individuals	PIK Rate
	(1)	(2)	(3)	(4)	(5)	(6)
2014	3,500	3,900	89.74%	8,700	9,500	91.58%
2015	3,900	4,300	90.70%	9,500	10,500	90.48%
2016	3,800	4,200	90.48%	9,000	10,000	90.00%
Total	11,200	12,400	90.32%	27,200	30,000	90.67%

Source: 2014-16 Consumer Expenditure Survey linked to various administrative records

Notes: This table presents individual-level and CU-level PIK rates for those interviewed between January and April of survey years 2015-2017 in the 1st and 4th wave with an unambiguous state indicator. The rates are based off values rounded in accordance with Census disclosure rules. All results were approved for release by the Census Bureau's Disclosure Review Board, authorization numbers CBDRB-FY23-0184 and CBDRB-FY2023-CES005-016

Table A3. Publicly Available Aggregates and their Sources

Survey Income Source	Public Aggregate Available?	Aggregate Source
	Panel A. Modified AGI	
Wages and Salary	Yes	NIPA (Wages and salary)
Self-Employment Income	Yes	NIPA (Proprietors' income, non-farm and farm)
Interest and Dividends	Yes	NIPA (Personal interest income, personal dividend income)
Rental Income	Yes	NIPA (Rental income)
Royalties	Yes	
Retirement and Pensions	Yes	NIPA (Defined contribution plans, Private pension plans, Federal civilian pension plans, Federal military pensions, State and local employee retirement)
Other REGULAR Income	Yes	U.S. Department of Veterans Affairs (Veteran's Benefits). McLaren, Baldwin, and Boden 2018 (Worker's Compensation). DOL (Unemployment Insurance), Office of Child Support Enforcement (Child Support), IRS SOI (Alimony)
Other NON-RENTAL Income	No	Weighted Survey Total
Total		
With Aggregate	Yes	
	Panel B. Select Government Programs	
OASDI + Railroad Retirement	Yes	SSA
SSI	Yes	SSA
Welfare	Yes	Department of Health & Human Services (AFDC/TANF), NIPA (General Assistance)
SNAP	Yes	Department of Agriculture
Housing Assistance, Dollars	Yes	CID (only for 2016)
Housing Assistance, Counts.	Yes	CID (only for 2016)
Total		
With Aggregate	Yes	
	Panel C. Tax Simulations	
AGI	Yes	SOI Line-Item Totals
Federal Tax	Yes	SOI Line-Item Totals
State Tax	Yes	Census Bureau Survey of State Governments
FICA	Yes	SSA for payroll tax liabilities
SECA	Yes	SSA for payroll tax liabilities
Taxable Income	Yes	SOI Line-Item Totals
Federal Tax before Credits	Yes	SOI Line-Item Totals
EITC	Yes	SOI Line-Item Totals
CTC (Refundable)	Yes	SOI Line-Item Totals
ACTC (Non-Refundable)	Yes	SOI Line-Item Totals

Notes: This table outlines the different sources employed to establish our aggregate benchmarks, against which we assess our weighted totals. We combine the methodologies of Meyer, Mok, and Sullivan (2015) and Rothbaum (2015) to create publicly available aggregates. Where applicable, we remove income received by the institutionalized, those living overseas, and military personnel. We gather data from IRS SOI line-item totals for CE specific miscellaneous income sources that cannot be found in the three papers above. We do not apply any adjustments to these values. To account for the Other NON-RENTAL Income category, for which we lack a publicly available aggregate, we utilize the weighted survey total based on the original CE survey weights.

Table A4. Comparisons of Weighted Survey Income Values to Public Aggregates (millions)

Income Source	CE Survey (1)	Aggregate (2)	Reporting Rate (3)
Panel A. Modified AGI			
Wage and Salary Income			
2014	6,756,000	7,436,064	90.85%
2015	7,171,000	7,817,057	91.74%
2016	7,521,000	8,046,884	93.46%
Self-Employment Income			
2014	536,700	1,200,979	44.69%
2015	855,800	1,138,178	75.19%
2016	564,700	1,133,042	49.84%
Retirement Pensions			
2014	304,200	1,086,956	27.99%
2015	296,900	1,138,936	26.07%
2016	314,400	1,154,629	27.23%
Interest and Dividends			
2014	108,000	946,172	11.41%
2015	118,700	1,044,846	11.36%
2016	108,100	1,099,004	9.84%
Rental Income and Royalties			
2014	91,500	160,833	56.89%
2015	118,000	151,419	77.93%
2016	111,400	148,419	75.06%
Other Regular Income (VA + WC + UI + Child Support + Alimony)			
2014	65,280	180,935	36.08%
2015	56,840	183,985	30.89%
2016	50,160	189,210	26.51%
Modified AGI			
2014	7,883,000	11,034,494	71.44%
2015	8,646,000	11,505,694	75.15%
2016	8,708,000	11,811,756	73.72%
Panel B. Select Government Programs			
OASDI			
2014	687,900	816,236	84.28%
2015	746,700	852,789	87.56%
2016	796,600	877,108	90.82%
SSI			
2014	29,360	54,612	53.76%
2015	29,690	55,482	53.51%
2016	22,110	55,281	40.00%
Welfare (TANF and GA)			
2014	5,506	25,810	21.33%
2015	3,407	27,111	12.57%
2016	2,445	27,552	8.87%
SNAP			
2014	38,150	69,511	54.88%
2015	39,300	68,841	57.09%
2016	35,100	65,552	53.54%
Housing Assistance, Dollars*			
2014	27,870		
2015	32,900		
2016	30,550	37,436	81.61%
Housing Assistance, Counts*			
2014	5,098,000		
2015	5,802,000		
2016	5,193,000	5,100,000	101.82%

Notes: This table presents the weighted totals and reporting rates of survey income components for the sample of 11,200 CUs interviewed in the 1st and 4th quarters. The weights are calculated per the methods in Data Appendix A.

Table A5a. Comparisons of Weighted Blended Income Values to Public Aggregates (mil.)

Income Source	CE Blended (1)	Aggregate (2)	Reporting Rate (3)
	Panel A. Modified AGI		
Wage and Salary (at individual level) *			
2014	7,581,000	7,436,064	101.95%
2015	7,989,000	7,817,057	102.20%
2016	8,426,000	8,046,884	104.71%
Wage and Salary (at tax unit level) **			
2014	7,503,000	7,436,064	100.90%
2015	7,948,000	7,817,057	101.68%
2016	8,384,000	8,046,884	104.19%
Self-Employment Income			
2014		1,200,979	
2015		1,138,178	
2016		1,133,042	
Retirement Pensions ***			
2014	1,098,000	1,086,956	101.02%
2015	1,127,000	1,138,936	98.95%
2016	1,220,000	1,154,629	105.66%
Interest and Dividends			
2014		946,172	
2015		1,044,846	
2016		1,099,004	
Rental Income and Royalties			
2014		160,833	
2015		151,419	
2016		148,419	
Other Regular Income (VA + WC + UI + Child Support + Alimony) ****			
2014	110,200	180,935	60.91%
2015	101,700	183,985	55.28%
2016	95,170	189,210	50.30%
Modified AGI			
2014	10,230,000	11,034,494	92.71%
2015	11,300,000	11,505,694	98.21%
2016	11,630,000	11,811,756	98.46%

*For the purposes of generating a weighted total, we blend wage and salary at the individual level using the following formula: max $\{Wages_{W2,}, Wages_{svy}\}$. This method reflects the blending procedure that we have for non-filers. Thus, when comparing the weighted total for this given category to that of its corresponding aggregate, it is important to note that it does not directly indicate how well our blending procedures work for the wage and salary of all individuals.

**We blend at the TU level using the following formula: max$\{Wages_{W2}, Wages_{1040}, Wages_{svy}\}$. This method reflects the blending procedure that we have for filers. However, unlike for non-filers, we do not blend wage and salary income directly for filers. Rather, we do so indirectly through the blending procedure for Modified AGI. To generating a weighted total, we blend directly here. As a result, our actual weighted total for wage and salary can be thought of as the sum of a non-random subset of wage and salary income blended at the individual level (non-filers) and wage and salary income blended at the TU level (filers).

***Like wage and salary income, we blend retirement pensions for filers through Modified AGI. However, for non-filers we blend retirement pensions at the component level using administrative data from Form 1099-R. Our weighted total is created using the maximum of administrative and survey retirement pensions at the CU level.

****For our administrative input in our blending procedure for other regular income, we only use VA benefits as we lack other administrative sources.

Table A5b. Comparisons of Weighted Blended Income Values to Public Aggregates (mil.)

Income Source	CE Blended (1)	Aggregate (2)	Reporting Rate (3)
Panel B. Select Government Programs			
OASDI			
2014	773,700	816,236	94.79%
2015	842,800	852,789	98.83%
2016	883,400	877,108	100.72%
SSI			
2014	46,660	54,612	85.44%
2015*	57,380	55,482	103.42%
2016	46,880	55,281	84.80%
Welfare (TANF and GA)			
2014		25,810	
2015		27,111	
2016		27,552	
SNAP**			
2014		69,511	
2015		68,841	
2016		65,552	
Housing Assistance, Dollars			
2014	55,700		
2015	60,810		
2016	55,530	37,436	148.33%
Housing Assistance, Counts			
2014	7,267,000		
2015	8,078,000		
2016	7,343,000	5,100,000	143.98%

*In reference year 2015, we find that there is a strong positive correlation between the original CE survey weights and SSI receipt amounts. This leads to a slightly elevated weighted total relative to 2014 and 2016.
**Our blended SNAP weighted total is the survey SNAP weighted total.

Table A5c. Comparisons of Weighted Blended Tax Outputs to Public Aggregates (mil.)

Income Source	CE Blended (1)	Aggregate (2)	Reporting Rate (3)
Panel C. TAXSIM27 Outputs			
AGI			
2014	9,349,000	9,771,035	95.68%
2015	10,010,000	10,210,310	98.04%
2016	10,650,000	10,225,938	104.15%
Federal Income Tax Liability*			
2014	1,055,000	1,262,872	83.54%
2015	1,259,000	1,345,734	93.55%
2016	1,348,000	1,339,996	100.60%
State Income Tax Liability*			
2014	282,800	349,827	80.84%
2015	306,700	383,712	79.93%
2016	320,900	384,307	83.50%
Taxable Income			
2014	6,167,000	6,997,856	88.13%
2015	6,819,000	7,350,296	92.77%
2016	7,191,000	7,330,109	98.10%
Federal Income Tax b/ Credits			
2014	1,136,000	1,367,933	83.05%
2015	1,327,000	1,449,892	91.52%
2016	1,413,000	1,440,430	98.10%
EITC			
2014	72,700	68,339	106.38%
2015	69,200	68,525	100.99%
2016	67,510	66,723	101.18%
CTC			
2014	31,500	27,202	115.80%
2015	31,360	27,100	115.72%
2016	31,150	26,800	116.23%
ACTC			
2014	20,620	27,063	76.19%
2015	19,380	26,590	72.88%
2016	20,340	25,373	80.16%
CTC + ACTC			
2014	52,090	54,264	95.99%
2015	50,670	53,690	94.37%
2016	51,400	52,174	98.52%
FICA			
2014	439,300	461,906	95.11%
2015	454,000	489,194	92.81%
2016	475,000	516,342	91.99%
SECA			
2014	66,400	62,666	105.96%
2015	85,480	60,739	140.73%
2016	66,490	60,296	110.27%
FICA + SECA			
2014	505,700	524,573	96.40%
2015	539,480	549,933	98.10%
2016	541,490	576,638	93.90%

*State and federal income tax liabilities are understated for most years. Since AGI and taxable income are relatively close to their respective benchmarks, the likely source of the difference comes from errors in estimating deductions and exemptions.

Notes: This table presents the weighted totals and reporting rates of blended income components for the sample of 11,200 CUs interviewed in the 1st and 4th quarters. The weights are calculated using the methodology described in Data Appendix A. We lack several income categories because we lack the corresponding component-level administrative data and thus include these income components through modified AGI. As a result, we do not have blended CU or individual values for the categories. All results were approved for release by the Census Bureau's Disclosure Review Board, authorization numbers CBDRB-FY23-0184 and CBDRB-FY2023-CES005-016.

Table A6. Poverty Rates for Those in All States

Demographic Subgroup	Survey Income with SNAP and Housing Assistance* (1)	Blended Income with SNAP and Housing Assistance (2)	Consumption (3)
Deep Poverty (0.5x Poverty Line)			
Full Sample	0.053	0.019	0.018
Single Adult w/o Children	0.130	-	0.038
Single Adult w/ Children	0.192	0.061	0.020
Multiple Adults w/o Children	0.042	-	0.012
Multiple Adults w/ Children	0.034	-	0.021
CUs w/ Elderly (65+)	0.042	0.011	0.013
Under 18	0.061	0.019	0.023
Aged 18 to 54	0.052	0.024	0.021
Aged 55 to 64	0.055	-	0.010
Aged 65 or older	0.040	0.013	0.009
Poverty (1x Poverty Line)			
Full Sample	0.130	0.073	0.179
Single Adult w/o Children	0.235	0.176	0.133
Single Adult w/ Children	0.304	0.215	0.299
Multiple Adults w/o Children	0.103	0.053	0.137
Multiple Adults w/ Children	0.107	0.059	0.214
CUs w/ Elderly (65+)	0.127	0.054	0.147
Under 18	0.144	0.086	0.239
Aged 18 to 54	0.128	0.078	0.193
Aged 55 to 64	0.125	0.052	0.112
Aged 65 or older	0.116	0.055	0.105
Near Poverty (1.5x Poverty Line)			
Full Sample	0.289	0.184	0.442
Single Adult w/o Children	0.396	0.304	0.345
Single Adult w/ Children	0.621	0.511	0.628
Multiple Adults w/o Children	0.192	0.118	0.365
Multiple Adults w/ Children	0.278	0.179	0.521
CUs w/ Elderly (65+)	0.301	0.150	0.369
Under 18	0.347	0.247	0.549
Aged 18 to 54	0.269	0.182	0.465
Aged 55 to 64	0.247	0.131	0.337
Aged 65 or older	0.298	0.142	0.302
Twice Poverty (2x Poverty Line)			
Full Sample	0.430	0.311	0.647
Single Adult w/o Children	0.504	0.423	0.545
Single Adult w/ Children	0.746	0.705	0.797
Multiple Adults w/o Children	0.284	0.198	0.559
Multiple Adults w/ Children	0.437	0.330	0.738
CUs w/ Elderly (65+)	0.480	0.274	0.575
Under 18	0.501	0.405	0.759
Aged 18 to 54	0.401	0.308	0.672
Aged 55 to 64	0.351	0.222	0.535
Aged 65 or older	0.484	0.258	0.501

*The survey income concept does not include housing assistance at deep poverty and near poverty.

Notes: This table presents the poverty rates for our main sample of CUs that include at least one member with a PIK, have an unambiguous state indicator, and are interviewed in the 1st and 4th waves between January-April of 2015-2017. The poverty line for 2016 is equal to the weighted average of the SPM thresholds by housing tenure. For other years 2014 and 2015, thresholds are deflated using the PCEPI. All results were approved for release by the Census Bureau's Disclosure Review Board, authorization numbers CBDRB-FY23-0184 and CBDRB-FY2023-CES005-016.

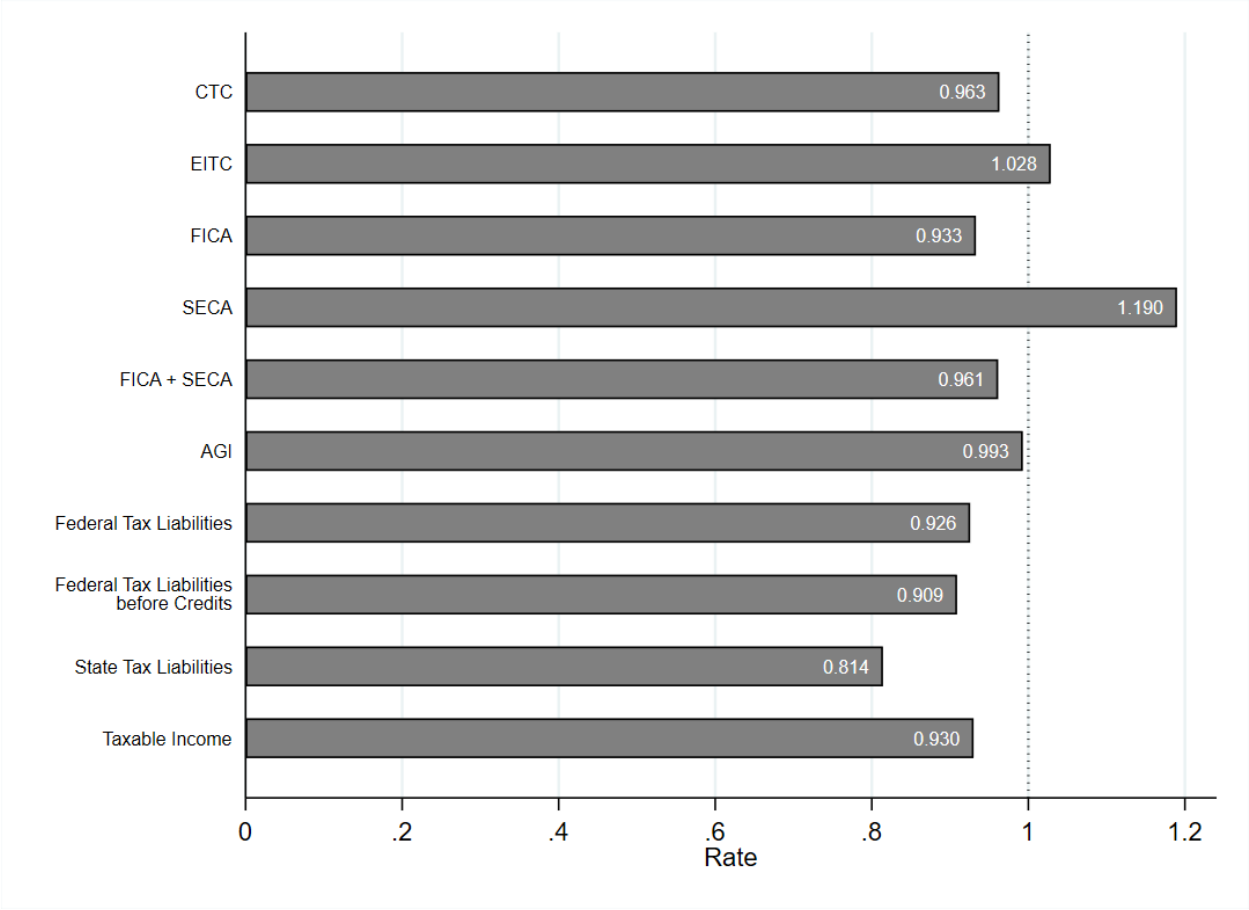
Table A7. Poverty Rates for Those in States with Administrative SNAP Data

Demographic Subgroup	Survey Income with SNAP and Housing Assistance* (1)	Blended Income with SNAP and Housing Assistance (2)	Consumption (3)
Deep Poverty (0.5x Poverty Line)			
Full Sample	0.049	0.018	0.014
Single Adult w/o Children	0.136	-	0.055
Single Adult w/ Children	0.151	0.066	-
Multiple Adults w/o Children	0.048	-	0.014
Multiple Adults w/ Children	0.027	-	0.011
CUs w/ Elderly (65+)	0.039	0.007	0.011
Under 18	0.048	0.017	0.012
Aged 18 to 54	0.051	-	0.021
Aged 55 to 64	0.052	-	-
Aged 65 or older	0.040	0.009	0.008
Poverty (1x Poverty Line)			
Full Sample	0.126	0.064	0.163
Single Adult w/o Children	0.264	0.185	0.128
Single Adult w/ Children	0.240	0.168	0.228
Multiple Adults w/o Children	0.115	0.054	0.133
Multiple Adults w/ Children	0.099	0.048	0.200
CUs w/ Elderly (65+)	0.118	0.043	0.131
Under 18	0.126	0.068	0.219
Aged 18 to 54	0.130	0.074	0.183
Aged 55 to 64	0.127	0.049	0.101
Aged 65 or older	0.112	0.043	0.085
Near Poverty (1.5x Poverty Line)			
Full Sample	0.289	0.167	0.428
Single Adult w/o Children	0.433	0.310	0.363
Single Adult w/ Children	0.567	0.401	0.583
Multiple Adults w/o Children	0.209	-	0.367
Multiple Adults w/ Children	0.274	0.164	0.506
CUs w/ Elderly (65+)	0.297	0.117	0.347
Under 18	0.328	0.206	0.526
Aged 18 to 54	0.283	0.180	0.466
Aged 55 to 64	0.243	0.116	0.327
Aged 65 or older	0.291	0.121	0.270
Twice Poverty (2x Poverty Line)			
Full Sample	0.430	0.305	0.636
Single Adult w/o Children	0.540	0.442	0.565
Single Adult w/ Children	0.701	0.665	0.781
Multiple Adults w/o Children	0.297	0.203	0.570
Multiple Adults w/ Children	0.428	0.324	0.718
CUs w/ Elderly (65+)	0.482	0.262	0.559
Under 18	0.481	0.386	0.738
Aged 18 to 54	0.409	0.315	0.672
Aged 55 to 64	0.352	0.211	0.547
Aged 65 or older	0.486	0.241	0.469

*The survey income concept does not include housing assistance at deep poverty and near poverty.

Notes: This table displays the poverty rates for the sample described in Table A6 (previous table) with the additional restriction of CUs in the 22 states with administrative SNAP data. The poverty line for reference year 2016 is derived from the weighted average of the SPM thresholds by housing tenure. For other years 2014 and 2015, thresholds are deflated using the PCEPI. All results were approved for release by the Census Bureau's Disclosure Review Board, authorization numbers CBDRB-FY23-0184 and CBDRB-FY2023-CES005-016.

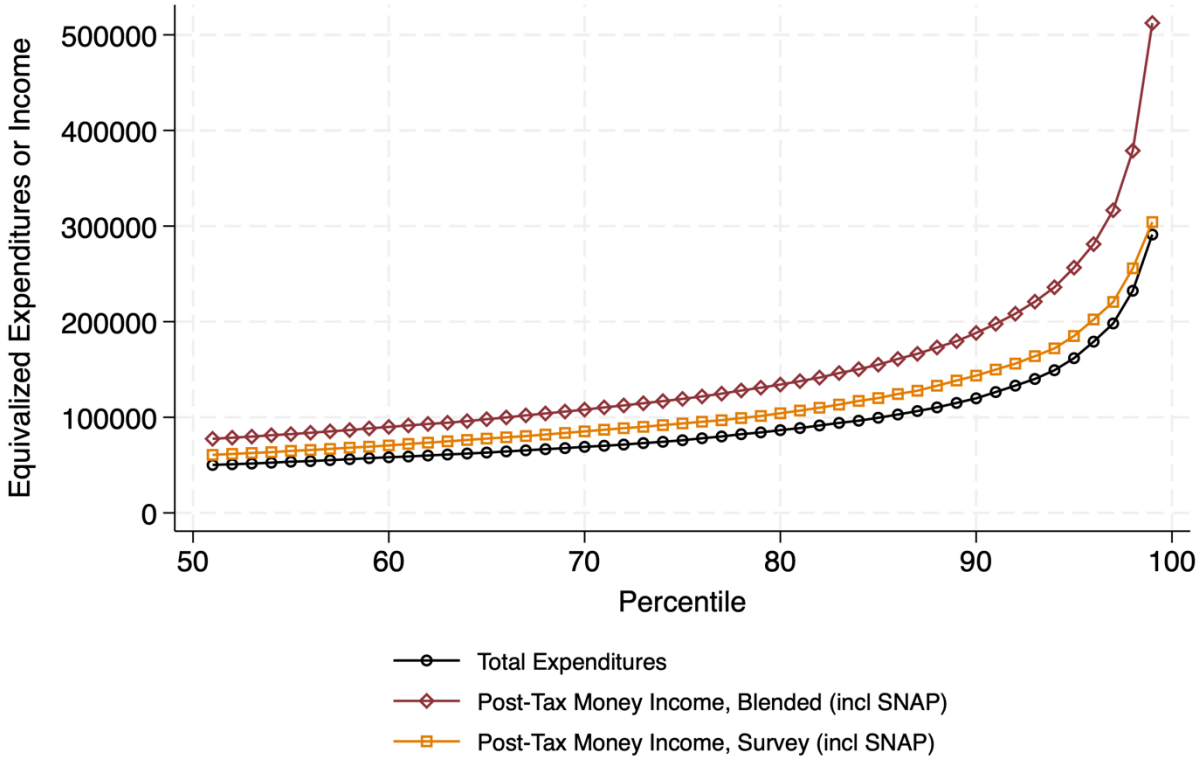
Figure A1. Average Reporting Rates of Tax Estimates Created using Blended Inputs



Source: 2014-16 Consumer Expenditure Survey linked to various administrative records, IRS SOI totals

Notes: This figure presents the average of yearly reporting rates for tax estimates simulated using survey and administrative data and outputted by TAXSIM. We compare the weighted totals of survey and blended income components to publicly available aggregates. A description of the weighting methodology can be found in Data Appendix A and the sources used to construct publicly available aggregates are listed in Appendix Table A3. The underlying numbers can be found in Appendix Table A5c. Approved for release by the Census Bureau's Disclosure Review Board, authorization numbers CBDRB-FY23-0184 and CBDRB-FY2023-CES005-016.

Figure A2. Distribution of Resource Measures, 51st-99th Percentiles

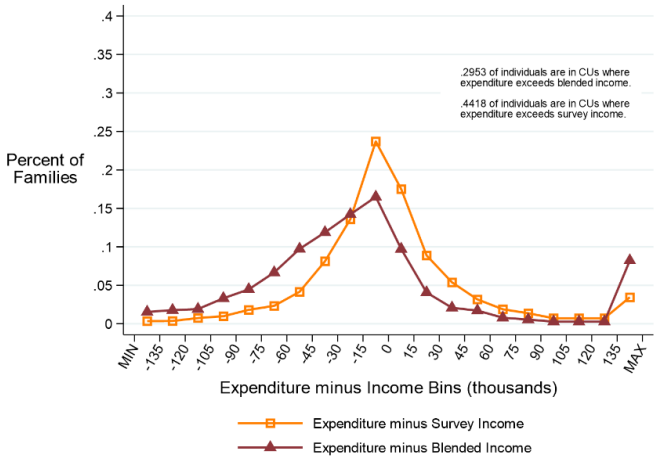


Source: 2014-16 Consumer Expenditure Survey linked to various administrative records

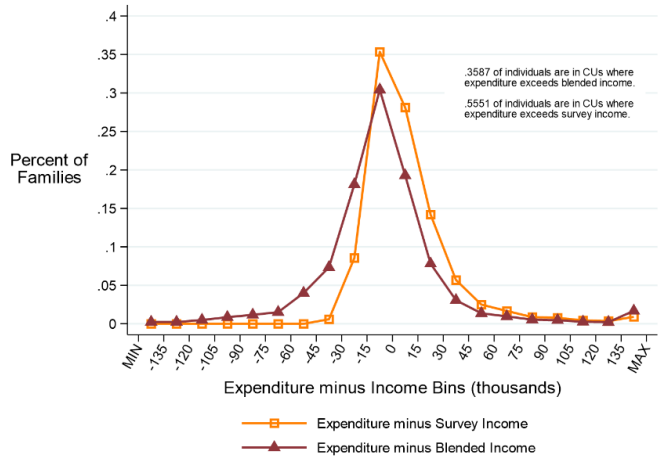
Notes: This figure shows the 51st through 99th percentiles for three resource measures – blended income, survey income, and total expenditures – pooled across the three years in our sample. All values are expressed in 2016 dollars and adjusted for differences in family size using the three-parameter SPM equivalence scale and normalized to a two-adult and two-child family. Income is after tax and includes SNAP. Both income and expenditures exclude housing assistance. Approved for release by the Census Bureau's Disclosure Review Board, authorization numbers CBDRB-FY23-0184 and CBDRB-FY2023-CES005-016.

Figure A3. Distribution of Expenditure minus Income, Subsamples

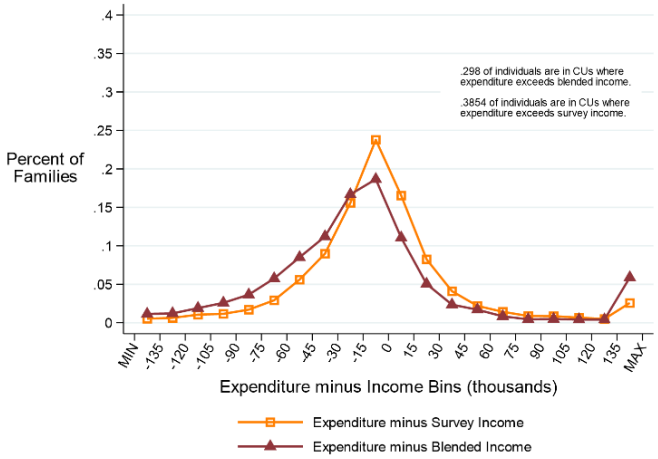
a) Elderly Individuals (65+)



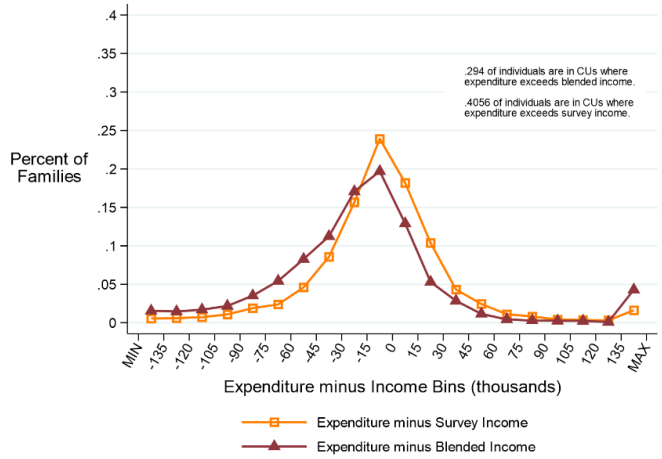
b) Individuals in CUs with Survey Income less then $50,000



c) Individuals in States with Administrative SNAP Data



d) Individuals in CUs interviewed in the 4th Wave.



Source: 2014-16 Consumer Expenditure Survey linked to various administrative records
Notes: We partition the data into $15,000 bins and position markers at the midpoints of bins. All values are expressed in 2016 dollars, adjusted for differences in family size using the three-parameter SPM equivalence scale, and normalized to a two-adult, two-child family. Income is after tax and includes SNAP. Approved for release by the Census Bureau's Disclosure Review Board, authorization numbers CBDRB-FY23-0184 and CBDRB-FY2023-CES005-016.

Data Appendix

Appendix A. Weighting Methodology

This section details the methodology that we use to generate inverse probability weights for PIKed individuals in either 1st and 4th or only 4th interviews with unambiguous state indicators.[1] We begin by running a probit model over our sample to predict the likelihood that a CU would have at least one individual attached to a PIK based on available survey and income and demographic information. We then multiply the existing CE survey weights with the inverse of the probability that it contains someone that is PIKed.

Our probit model controls for the following factors: age, family type, education level, race, Hispanic origin, sex, an interaction term for whether an individual is Hispanic and in a household making less than $20,000 per year, work status, marital status, pre-tax money income, household expenditures, urbanicity, residence in public and subsidized housing, indicator variables for SNAP, OASDI, SSI, and TANF receipt, and the share of the CU that reports coverage by any medical and private insurance. All individual-level demographic information such as age and race is taken from the reference person only.

We then re-weight the PIKed observations to match CPS ASEC population totals by full interactions of age-family type and education of the head for the reference year categories. Our age-family type categories are divided into four non-elderly categories (multiple adults with children, multiple adults without children, single adult with children, single adult without children) and elderly (65+). The education of head categories are divided into high school or less, some college, and four-year college or more.

[1] We have a different set of weights depending on whether we restrict to 1st and 4th interviews or just 4th interviews.

Appendix B. Measuring Consumption and Expenditures in the CE

This section provides additional details on expenditures and consumption and the adjustments made to them.

Total Expenditure: This resource measure includes all expenditures reported in the CE Interview Survey except for contributions to retirement, pensions, and Social Security because they are not considered to be new economic activity.[2]

Total Consumption: Total consumption is the aggregate spending on goods and services by households, excluding out-of-pocket health care expenses, education, and payments to retirement accounts, pension plans, and Social Security. Housing and vehicle expenditures are converted to service flows, which is the value of the services provided by these goods, rather than the actual price paid for them. For homeowners, this is done by subtracting spending on mortgage interest, property taxes, maintenance, repairs, insurance, and other expenses, and then adding the imputed rental equivalent of the home. The rental value for those in public or subsidized housing is imputed using a specific procedure detailed below. For vehicle owners, spending on recent purchases of new and used vehicles as well as vehicle finance charges is subtracted, and then the service flow value of all vehicles owned by the family is added.

Estimating Vehicle Service Flows

As a part of constructing our measure of consumption, we replace the purchase price of vehicles and vehicle maintenance costs with the service flow value from owned vehicles. Our methodology of imputing vehicle service flows follows the approach used in Han, Meyer, and Sullivan (2021) and Meyer, Murphy, and Sullivan (2022).

We begin with detailed expenditure data for owned vehicles from the 1980-2020 CE. We use one of three ways to determine a current market price for each of the 1.6 million vehicles in the data. In the case of vehicles purchased within the twelve-month period preceding the interview, and for which a purchase price has been reported (referred to as the estimation sample), we use the reported price as the current market value. On the other hand, for vehicles purchased over twelve months before the interview and for which a purchase price has been reported (15% of all vehicles), we derive the current market value through a function involving both the reported purchase price and an estimated depreciation rate, detailed below.

We impute the current market price for the remaining 72 percent of vehicles, as the purchase price is not reported in the survey. Using the estimation sample, we regress the log real purchase price, $\log(y_i)$, on a cubic in vehicle age, vehicle characteristics, family characteristics, and make-model year fixed effects. The vehicle characteristics include indicators for whether the vehicle has automatic transmission, power brakes, power steering, air conditioning, a diesel engine, a sunroof, four-wheel drive, or is turbo charged. Family characteristics include log real expenditures (excluding vehicles and health), family size, region, and the age and education of the family head. Coefficient estimates from this regression are then used to calculate a predicted log real purchase price for vehicle i, $\left(x_i\hat{\beta}\right)$. The predicted current market value for each vehicle

[2] These values correspond to UCC 800910, 800920, 80093, 800932, and 800940.

without a reported purchase price is then equal to $\hat{\alpha} * exp(x_i\hat{\beta})$ where $\hat{\alpha}$ is the coefficient on $exp(x_i\hat{\beta})$ in a regression of y_i on $exp(x_i\hat{\beta})$ without a constant term.

To estimate a depreciation rate for vehicles, we compare prices across vehicles of different age, but with the same make, model, and year. In particular, from the estimation sample we construct a subsample of vehicles that are in a make-model-year cell with at least two vehicles that are not the same age. Using this sample, we regress the log real purchase price of the vehicle-on-vehicle age and make-model-year fixed effects. From the coefficient on vehicle age, β, we calculate the depreciation rate δ, where $\delta = 1 - exp(\beta)$. The service flow is then the product of this depreciation rate and the current market price. If the vehicle has a reported purchase price but was not purchased within 12 months of the interview, we calculate the service flow as: $(real\ reported\ purchase\ price) * \delta * (1-\delta)^t$, where t is the number of years since the car was purchased.

We validate our procedure for predicting the current market value of vehicles for those observations where we do not have a purchase price by comparing the predicted values to published values in National Automobile Dealers Association (NADA) guides. For a given year of the CE we take a random sample of 100 vehicles for which a purchase price was not observed. We then find the average retail price of the vehicle reported in the NADA Official Used Car Guide, using observable vehicle characteristics including make, model, year, number of cylinders, and number of doors. In cases where a unique match is not found in the NADA guide (for example, there might be multiple sub-models listed in the NADA guide), we use the midpoint of the range of prices for the vehicles that match the description of the vehicle from the CE. For the sample of vehicles randomly drawn from the 2000 CE, the correlation between our imputed price and the 2000 NADA price was 0.88. Similarly, for a sample of 100 cars with a reported purchase price, the correlation between the reported price and the NADA price was 0.91.

Estimating a Rental Equivalent for Families Living in Government or Subsidized Housing

We impute a rental equivalent for families in the CE living in government or subsidized housing using reported information on their living unit, including the number of rooms, bedrooms, and bathrooms and the presence of appliances such as a microwave, disposal, refrigerator, washer, and dryer. Specifically, for renters who are not in public or subsidized housing, we regress log rent on the CE housing characteristics mentioned above as well as a number of geographic identifiers including state, region, urbanicity, and SMSA status, and interactions of a nonlinear time trend with appliances (to account for changes over time in their price and quality). We then use the estimated coefficients to predict rent for the sample of families that do not report full rent because they reside in public or subsidized housing. We do not adjust for the lower quality of public housing in dimensions we do not directly observe. Evidence from the Panel Study of Income Dynamics indicates that the average reported rental equivalent of public or subsidized housing is just under the predicted 40th percentile for these units, using parameters estimated from those outside public or subsidized housing.

Appendix C. Constructing a CE Measure of Survey-Only Income

The following table contains a description of all the income components used to create our measure of survey-only income.

Income Category	Variable	Observation Level	CE Question
Salary and Wages	salaryxm	Individual	Did you/(NAME) receive any wages, salary, tips, bonuses, or commissions? How much did You/(Name) receive before taxes?
Self-Employment	sempfrmm	Individual	Did You/(Name) receive any self-employment income or have a loss? (Report income from own businesses (farm or non-farm) including proprietorships and partnerships.)
Retirement Pensions	retsurvm	CU	Did you or any member of your household receive any retirement, survivor, or disability pensions? What was the amount?
Social Security	socrrxm	Individual	What was the amount of the last Social Security or Railroad Retirement payment received?
SSI	ssixm	Individual	Did you receive any - Supplemental Security Income (SSI) payments? What was the amount?
Interest and Dividends	intrdvxm	CU	Did you or any member of your household receive any interest or dividends? Report even small amounts credited to an account. What was the total amount earned by all household members?
Rental Income	othregxm	CU	Did you or any member of your household receive any net rental income or a loss? What was the total amount earned by all household members? Net rental income is the total amount after expenses.
Royalties	royestxm	CU	Did you or any member of your household receive any royalty income or income from estates and trusts? What was the amount?
Other Regular Income	othregxm	CU	Did you or any member of your household receive income on a REGULAR basis from any other source such as Veteran's Administration (VA) payments, unemployment compensation, child support, or alimony? What was the amount from all sources? (Do not include lump sum payments such as money from an inheritance or sale of a home.)
Other Non-Rental Income	othrincm	CU	Did you or any member of your household receive any other money income, including money received from cash scholarship and fellowship, stipends not based on working, or from the care of foster children, not already reported? What was the total amount received by all household members?
SNAP	jfs_amtm	CU	Did anyone in this household receive Food Stamps or a Food Stamp benefit card? Include government benefits from the Supplemental Nutritional Assistance Program (SNAP). Do NOT include WIC of the National School Lunch Program.
Welfare	welfarem	CU	Did you/you or any members of this household, including any children, receive any welfare payments or cash assistance from the state or local welfare office? What was the amount for the PAST 12 MONTHS? Please include even if only for one month. Do NOT include benefits from food, energy, or rental assistance programs.
Housing Assistance[3]	cutenure publhous govtcost	CU	The survey question in the CE the following questions which we use to identify CUs that receive housing assistance: 1. Housing Tenure (CUTENURE) 2. Is this house in a public housing project, that is, is it owned by a local housing authority or other local public agency (yes/no)? (PUBLHOUS) 3. Are your housing costs lower because the Federal, State, or local government is paying part of the cost (yes/no)? (GOVTCOST) We impute the value of Housing Assistance using the following formula: IF cutenure GT 3 AND (publhous=1 OR govtcost=1) THEN quarterly rent = MAX(SUM(OF rendwepq rendwecq),mo_rent40*3)

[3] We exclude the value of housing assistance in income when comparing income to expenditures.

Appendix D. Methodology for Constructing Blended Income

Our goal of creating a more comprehensive measure of post-tax, post-transfer income involves blending together survey and administrative data sources. Our blending methodology uses Adjusted Gross Income (AGI) as a starting point for constructing a blended measure of pre-tax money income for individuals to whom we can attach to a 1040 (filers). We subtract several income components from AGI to create a version of AGI that we call modified AGI (sometimes referred to as *AGI**). For those to whom we cannot attach to a 1040 (non-filers), we use individual income components to construct something that is equivalent to modified AGI, blending in administrative data when available. As a part of our blending methodology, we divide consumer units (CUs) into three categories: "filer" CUs are CUs containing only filers, "non-filer" CUs are CUs containing only non-filers, and "mixed" CUs are CUs containing both filers and non-filers.

Blending Procedure for Filer CUs

For filers, the CU in the CE is equivalent to the Tax Unit (TU) in the 1040. This constitutes our simplest case as we can use 1040 values to account for all individuals within the CU.

1. We calculate our **survey modified AGI value** by subtracting survey SSI, public assistance/welfare, and OASDI from the CE's definition of pre-tax money income:

   $$AGI^*_{survey,CU} = pre\ tax\ money_{survey,CU} - SSI_{survey,CU} - public\ assistance_{survey,CU} - OASDI_{survey,CU}$$

2. We then calculate our modified AGI value using the AGI listed on the 1040 as the base. We add deferred compensation, tax-exempt interest, administrative VA benefits, and subtract taxable OASDI to match our survey definition of modified OASDI. We call this value the **administrative 1040 modified AGI**:

   $$AGI^*_{1040} = AGI_{1040} + deferred\ compensation_{1040} + tax\ exempt\ interest_{1040} + VA_{admin} - taxable\ OASDI_{admin}$$

3. Next, we calculate a version of modified AGI exclusively using available administrative wage and salary data from IRS Form W-2, as well as retirement pension information extracted from IRS Form 1099-R. The sum of these two income components constitutes the closest approximation to the concept of administrative 1040 modified AGI without utilizing any inputs directly from Form 1040 itself. We call this the **administrative component modified AGI**:

   $$AGI^*_{admin\ comp,TU} = wages_{W2,TU} + retpen_{1099R,TU}$$

4. We then calculate our best measure of administrative modified AGI by taking the maximum value of our administrative 1040 modified AGI and the administrative component modified AGI.

   $$AGI^*_{admin,CU} = \max\{AGI^*_{admin\ comp,TU}\ ,\ AGI^*_{1040}\}$$

5. We then take the maximum of survey modified AGI and administrative modified AGI to create what we call **blended modified AGI**:

$$AGI^{*}_{blended,CU} = max\{AGI^{*}_{admin,CU}, AGI^{*}_{survey,CU}\}$$

6. To arrive at our blended measure of post-tax, post-transfer income, we add or subtract additional blended income components to blended modified AGI. These include OASDI, SSI, public assistance, SNAP, housing assistance, and net tax liabilities:

$$\begin{aligned} post\ tax,\ post\ transfer_{blended} \\ &= AGI^{*}_{blended} + OASDI_{blended} + SSI_{blended} + public\ assistance_{survey} \\ &+ SNAP_{blended} + housing\ assistance_{blended} \\ &- net\ tax\ liabilities_{blended} \end{aligned}$$

The blending procedures of these different income components are detailed below. It should be noted that since we do not directly rely upon 1040 AGI to incorporate these components into our final measure of income, the blending methodology for these will stay constant over our three different CUs.

<u>OASDI</u>

To construct a blended measure of OASDI, we take the maximum of tax unit-level OASDI amounts from the 1040s, the sum of individual OASDI amounts from the MBR/PHUS, and the sum of individual OASDI amounts from the survey (within a CU):

$$OASDI_{blended,\,CU} = max\left\{ \begin{array}{c} OASDI_{1040,\,TU} \\ \sum_{i \in CU} \begin{cases} OASDI'_{MBR/PHUS,\,i}, & if\ impute = 0 \\ OASDI_{survey,\,i}, & if\ impute = 1\ or\ PIKed = 0 \end{cases} \end{array} \right.$$

Additionally, given that 1040 and survey OASDI values include Medicare premiums while PHUS/MBR values are net of Medicare premiums, we impute Medicare premiums for individuals that indicate OASDI receipt as follows:

$$OASDI'_{MBR/PHUS,\,i} = \begin{cases} OASDI_{MBR/PHUS,\,i} + 104, & if\ age \geq 65 \\ OASDI_{MBR/PHUS,\,i}, & otherwise \end{cases}$$

Here, $OASDI_{MBR/PHUS,\,i}$ refers to the raw administrative OASDI data from the MBR/PHUS and 104 corresponds to the basic premium paid in 2014.

<u>SSI</u>

To create a blended measure of SSI, we take the administrative SSI amount if the individual is PIKed and the survey SSI amount if the individual is un-PIKed:

$$SSI_{blended,i} = \begin{cases} SSI_{admin,i}, & if\ PIKed = 1 \\ SSI_{survey,i}, & if\ PIKed = 0 \end{cases}$$

Public Assistance (TANF)
As we do not have cleaned administrative TANF data, we use the survey component as a substitute.

SNAP
The blending procedure for SNAP varies depending on the sample under consideration. When conducting analyses for the entire sample, we use survey SNAP amounts. When conducting analyses on the subsample of 22 SNAP states, we rely exclusively on administrative SNAP values.

$$SNAP_{blended,i} = \begin{cases} SNAP_{admin,i}, & if\ sample = SNAP\ states \\ SNAP_{survey,i}, & if\ sample = All\ states \end{cases}$$

Housing Assistance
When blending housing assistance, we take the maximum of scaled assistance unit benefit amounts and housing assistance reported in the survey.

$$Hous\ Assist_{blended,CU} = \max\{\sum_{i \in CU} Hous\ Assist_{admin,i}, Hous\ Assist_{survey,CU}\}$$

Net tax liabilities
We obtain blended net tax liabilities after inputting administrative and survey income variables into TAXSIM 27 on the IRE.

Blending Procedure for Non-filer CUs

Since we lack 1040 information for everyone in non-filer CUs, we cannot use AGI as a baseline for our income calculations. As a result, we must make use of survey components, blending in administrative data when possible.

1. The initial step for our blending procedure involves creating blended income components when available. The blending methodology for the following income components is described below.

   Wages and Salaries
   To blend wage and salary income, we take the maximum of the sum of W-2 wages and salaries plus deferred compensation and survey wages and salaries at the individual level. We add deferred compensation to W-2 wages and salaries because the W-2 value is post-deduction whereas the survey value is pre-deduction.

   $$wages_{blended,i} = \max\{wages_{W2,i} + defcomp_{W2,i}, wages_{survey,i}\}$$

   Retirement, Survivor, and Disability income
   Retirement pensions are recorded at the individual level in 1099-Rs but are recorded at the CU level for survey data. To blend retirement pensions, we therefore take the maximum of

1) individual-level retirement pension values on 1099-Rs summed across all members of the CU and 2) survey-reported retirement pensions.

$$retpen_{blended,CU} = \max\{\sum_{i \in CU} retpen_{admin,i}, retpen_{survey,CU}\}$$

Other Regular Income / VA

We only have administrative VA benefits at the individual level, but VA along with other things deemed by the CE as "regular income" are recorded at the CU level for the survey. Thus, the blending procedure for other regular income/VA benefits also involves taking the maximum of individual-level VA benefits summed across all members of the CU and survey reported other regular income.

$$othreginc_{blended,CU} = \max\{\sum_{i \in CU} VA_{admin,i}, othreginc_{survey,CU}\}$$

2. To construct the blended equivalent of modified AGI for CUs that cannot be linked to 1040 data, we take the survey modified AGI value and substitute in the blended components where possible. As a result, to construct blended modified AGI for non-filers, we take the sum of total blended wages at the CU level, survey self-employment, blended retirement pensions, survey interest, dividends, and royalties, survey rental income, blended other regular income, and survey other non-rental income.

$$\begin{aligned} AGI^*_{blended} = & \sum_{i \in CU} wages_{blended,i} + self\ employment_{survey,CU} + retpen_{blended,CU} \\ & + int,\ div,\ roy_{survey,CU} + rental\ inc_{survey,CU} + oth\ reg\ inc_{blended,CU} \\ & + other\ inc_{survey,CU} \end{aligned}$$

3. We then add other blended income components and subtract net tax liabilities from blended modified AGI to create a post-tax, post-transfer measure of income.

$$\begin{aligned} & post\ tax,\ post\ transfer_{blended} \\ & = AGI^*_{blended} + OASDI_{blended} + SSI_{blended} + public\ assistance_{survey} \\ & + SNAP_{blended} + housing\ assistance_{blended} \\ & - net\ tax\ liabilities_{blended} \end{aligned}$$

The blending methodologies are the same as in Step 6 of the blending procedures for filers.

Blending Procedure for Mixed CUs

The process of blending income for mixed CUs is more involved, as we have both filers and non-filers within a single consumer unit. This means that the consumer unit is not equivalent to the tax unit and thus any income categories derived from 1040 variables may not include the income earned by the non-filers in the CU. This implies that we must use other administrative data to blend income for non-filers in the Mixed CU.

1. We begin our blending procedure by calculating our best measure of admin modified AGI for only filers in the mixed CU. This involves taking the maximum of administrative component modified AGI and administrative 1040 modified AGI. The steps for creating initial admin 1040 modified AGI value and the admin component modified AGI value are identical to steps 2 and 3 in the blending procedure for filer CUs. The only difference is that we do this at the TU level where the TU is not equal to the CU.

$$AGI^*_{admin,TU} = \max\{AGI^*_{admin\ comp,TU}\ ,\ AGI^*_{1040,TU}\}$$

2. We then move on to calculating our administrative modified AGI equivalent for only non-filers in the mixed CU. This involves adding the sum of blended wages and salary, administrative VA benefits, and retirement benefits across all non-filers in a CU. While we technically use survey data here to blend wages and salary, we still refer to the modified AGI measure as "administrative". Wage and salary is the only income category subject to blending in this context due to the availability of both survey and administrative data recorded at the individual level. The blending process for wages and salary mirrors the methodology outlined earlier in step 1 of the blending procedure for non-filer CUs.

$$AGI^*_{admin,CU-TU} = \sum_{i\ \in CU-TU} wages_{blended,i} + \sum_{i\ \in CU-TU} VA_{admin,i} + \sum_{i\ \in CU-TU} retpen_{admin,i}$$

3. Having created an administrative modified AGI for both filers and non-filers, we add administrative modified AGI for filers and non-filers to create administrative modified AGI at the CU level.

$$AGI^*_{admin,CU} = AGI^*_{admin,TU} + AGI^*_{admin,CU-TU}$$

4. We then create blended modified AGI by taking the maximum of administrative modified AGI and survey modified AGI. The method for constructing survey modified AGI component is identical to Step 1 of the blending procedure for filers.

$$AGI^*_{blended,CU} = max\{AGI^*_{admin,CU}, AGI^*_{survey,CU}\}$$

5. We then add blended OASDI, SSI, SNAP, housing assistance, and survey public assistance to and subtract net tax liabilities from blended modified AGI to create a post-tax, post-transfer measure of income.

$$\begin{aligned} post\ tax,\ post\ transfer_{blended} &= AGI^*_{blended} + OASDI_{blended} + SSI_{blended} + public\ assistance_{survey} \\ &+ SNAP_{blended} + housing\ assistance_{blended} \\ &- net\ tax\ liabilities_{blended} \end{aligned}$$

The blending methodologies are the same as outlined in Step 6 of the blending procedure for filers.

Appendix E. Calculating Taxable OASDI

A challenge associated with blending survey and administrative Social Security (OASDI) data using AGI as a proxy is that AGI only contains the value of the taxable portion of OASDI while the survey concept includes total OASDI. To reconcile this mismatch, we create an adjusted version of AGI that excludes the value of taxable OASDI so that we may incorporate the value of blended OASDI separately. However, because we lack a separate variable for the taxable amount of OASDI on the 1040, we calculate it using the following formula:

$$taxable\ OASDI = min\left\{\begin{matrix} 0.85 * net\ OASDI \\ min\left\{\begin{matrix} 0.5 * net\ OASDI \\ max\left\{\begin{matrix} 0 \\ 0.5 * min\left\{\begin{matrix} (AGI' - K) \\ L - K \end{matrix}\right\} \end{matrix}\right\} \end{matrix}\right\} + max\left\{\begin{matrix} 0 \\ 0.85 * (AGI' - L) \end{matrix}\right\} \end{matrix}\right\}$$

where $net\ OASDI$ refers to OASDI benefits (net of repayments and deductions for work, not of Medicare deductions) and AGI' is a version of AGI less taxable OASDI benefits plus tax-exempt interest income plus exclusions and adjustments plus deductions.[4] K and L are two thresholds related to filing status that are used to calculate taxable OASDI and are defined as follows:

$$K = \begin{cases} \$32{,}000, & married\ filing\ jointly \\ \$0, & MFS\ but\ lived\ w/\ spouse \\ \$25{,}000, & otherwise \end{cases} \qquad L = K + \begin{cases} \$12{,}000, & married\ filing\ jountly \\ \$0, & MFS\ but\ lived\ w/\ spouse \\ \$9{,}000, & otherwise \end{cases}$$

Considering the recurrence of taxable OASDI on both sides of the equation, we solve the equation numerically. We set the initial value of taxable OASDI to \$0 and evaluate the RHS. If the value satisfied the convergence criterion that the difference between the LHS and the RHS is less than or equal to \$1, then we stop. Otherwise, the previous step is repeated until the value converges.

[4] Exclusions and adjustments include adoption benefits, foreign earned income, and certain American Samoa & Puerto Rico income. We do not have these values in the administrative data and assume that they are zero. Deductions include items such as student loan interest, tuition & fees, and domestic production activities. Again, we do not have these values in the administrative data and assume that they are zero.

Appendix F. Adjustments to Resource Measures

We adjust all our resource measures for inflation by using the Personal Consumption Expenditures Chain-Type (PCE) price index. More specifically, we take the average of monthly PCE index values for each reference year, standardizing to 2016 dollars.

Additionally, given that families will vary in the number of adults and children, we use the SPM three-parameter equivalence to adjust our resource measures, standardizing to a family with two adults and two children. The equivalence scale is as follows:

$number\ of\ adults^{0.5}$ for one and two adult families,
$[number\ of\ adults + 0.8 + 0.5(number\ of\ children - 1)]^{0.7}$ for single parent families, and
$[number\ of\ adults + 0.5 * number\ of\ children]^{0.7}$ for all other families.