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

December 27, 2024

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, with the new data having important impacts on the analyses. Our income measures relying on combined survey and administrative data compare much more closely than survey-based aggregates to benchmark totals from national accounts and other sources. The distributions of these combined, or "blended", income measures are also closer to those of expenditure measures. This resolution is especially true for the very bottom of the distribution, where prior research has revealed concerns about the under-reporting of survey income. We also see less evidence of under-reporting with the blended income measures, with fewer individuals having expenditures that exceed their 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. But over the entire distribution, household level income and expenditures are not made appreciably closer by improving the income measure.

* 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 an issue 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 and 2024 CNSTAT Fall Seminar 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. All of the authors declare that they have no competing interests or further disclosures.

1. Introduction and Motivation

The literature on the distribution of income is very long and deep. There is an especially large literature about the bottom of the distribution and on measures of poverty. There are also large literatures on the role of various government programs and income sources in contributing to the income of those with few resources and keeping them from being poor. Much of this work has focused on particularly disadvantaged groups, such as the elderly, disabled, and single parent families. A difficulty with these literatures that is increasingly recognized is that income sources, particularly for those at the bottom of the reported income distribution, are often mismeasured and most commonly under-reported. This situation has stimulated two approaches to address income under-reporting. One approach combines survey and administrative data on income to improve income measures. A second approach relies on expenditure data rather than income data to create measures of consumption. In this paper, we consider both approaches applied to the same individuals in the same dataset. This is the first research to directly compare measures of expenditures and consumption to income that has been corrected for a substantial part of misreporting. It builds on the need highlighted in NASEM (2024) to construct a dataset with both income and consumption to examine their joint distribution. As consumption is difficult to measure, the report suggests pilot studies to improve the CE survey data.[1]

The evidence on income under-reporting in the main Census Bureau household surveys has burgeoned in recent years. A line of papers has compared weighted survey totals of income sources to national income accounts and government payment totals. The most recent and comprehensive summaries of these aggregate comparisons can be found in Meyer, Mok and Sullivan (2015) and Rothbaum (2015). These papers find that large shares of social insurance and welfare payments, as well as self-employment income, retirement pensions, and many non-labor income sources are sharply under-reported on average. Significantly, these comparisons indicate that under-reporting for many income sources and surveys has increased over time.

A second approach to examining the misreporting of income compares individual microdata from surveys to government program records linked at the individual or family level. Results from such linkages are available for fewer income sources and years but provide a similar

[1] Among these pilot studies are 1) expanding NEWS to include more income sources, 2) measuring spending in commercial data to improve CE expenditure data, and 3) linking data to compare estimates of income and population characteristics in the tails of the distributions (NASEM 2024, p. 244).

picture to the aggregate analyses. In the case of the Current Population Survey, the source of official income and poverty statistics, Meyer and Mittag (2019) find that 43 percent of SNAP recipients, 63 percent of TANF or General Assistance Recipients and 36 percent of housing assistance recipients in New York State over the 2008-2011 period did not report receipt. Meyer, Wu, Stadnicki, and Langetieg (2022) find that 39 percent of UI recipient nationwide did not report receipt in 2010. Under-reporting is also apparent in the Survey of Income and Program Participation, which is generally thought to be the Census Bureau survey with the most complete income data (Gathright and Crabb 2014; Meyer, Mittag and Goerge, 2018). The former paper shows a pronounced increase over time in false negative reporting of Social Security retirement, disability, and SSI benefits. And beyond just government benefits being under-reported, about half of private pension recipients do not report receipt in our major household surveys (U.S. Census Bureau 2016; Bee and Mitchell 2017).

Recently, because of concerns about these survey responses, agencies and researchers have changed their strategies. The Social Security Administration stopped issuing after the 2014 editions two long-term publications – *Income of the Population 55 or Older* and *Income of the Aged Chartbook*. Researchers have begun to link administrative data to household surveys to replace misreported earnings, pension, and program data with administrative versions. This approach has been especially important in studies of older populations. These studies, particularly those incorporating tax records, have found that incomes of the elderly are often much higher than reported in the survey data alone, while the impact of SSA programs is often different as well (Nicholas and Wiseman 2009, 2010; Bee and Mitchell 2017; Dushi and Trenkamp 2021). Research on broader populations has found that the share of individuals below the poverty line and its multiples are much lower after replacing key income components with their administrative data equivalents (Meyer and Mittag 2017; Meyer and Wu 2022). This pattern is particularly true when examining the share below half the poverty line or even lower (Meyer et al. 2022), and the poverty reduction of many social insurance and welfare programs is misstated, most commonly understated (Meyer and Wu 2018).

A complementary approach to improving income measurement through linkage, and the one that we include in the current study, is to examine consumption. Consumption measures have the advantage of reflecting past savings and debt, the flow of services from owned houses and cars,

and accounting for the pronounced under-reporting of pension and transfer income.[2] These measurement issues are likely to be particularly important for older individuals given the pronounced under-reporting of pension payouts, as well as other groups that commonly receive social insurance or means-tested benefits (see Dushi et al. 2024).

In this paper, we examine the joint distribution of income and consumption-based measures of well-being. We are the first to conduct such analyses using linked expenditure and administrative income data, as stressed by NASEM (2024) in Recommendation 5-2, with the new data having important impacts on the analyses.[3] Specifically, this paper reflects the suggestions in Figure 5-1A (NASEM 2024, p. 234), as we blend tax data – along with other administrative program records – with the incomes from survey data that also include consumption. This is similar to the approaches put forward in prior work by the Comprehensive Income Dataset (CID) and National Experimental Wellbeing Statistics (NEWS) projects (Bee et al. 2023, Meyer and Wu 2024) and addresses the pilot studies of improving consumer expenditure data using blended data. While we examine the entire income and consumption distributions, we focus on the bottom percentiles as well as poverty measurement because it is particularly hard to estimate income and consumption for the bottom tail and government program benefits are more important at the bottom.

Our income measures relying on combined survey and administrative data compare much more closely than survey-based aggregates to benchmark totals from national accounts and other sources. The distributions of these combined, or "blended", income measures are also closer to those of expenditure measures. This is especially true for the very bottom of the distribution, where prior research has revealed concerns about the under-reporting of survey income. We also see less evidence of under-reporting with the blended income measures, with fewer individuals having expenditures that exceed their 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.

---

[2] We include pension payments in income as we are focusing on measures of contemporary well-being and want to step closer to expenditures and consumption. This is consistent with NASEM (2024), which states "For the purpose of most distributional analyses…the preferred income concept will include retirement income distributions" (p. 77).

[3] 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 paper proceeds as follows. Section 2 describes our survey and administrative data sources, while in Section 3 we describe our methods of combining survey and administrative data. Section 4 reports comparisons of our survey and blended data to administrative aggregates. Section 5 describes the distribution of expenditures and income. Section 6 reports poverty rates. Lastly, Section 7 provides a discussion, conclusions, and future directions.

2. Data: The CE Survey, Samples, Administrative Data, and Linkage

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 in turn linked to administrative tax records from the Internal Revenue Service (IRS) and administrative program records from various government agencies, which enable us to correct measurement errors in survey income components. We focus on data for calendar years 2014-2016, because of the availability of both survey and administrative data for those years.

The Consumer Expenditure Survey and Sample Definition

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. The CE is structured as a rotating panel that includes about 7,500 families each quarter. Respondents report both their spending on a large number of expenditure categories as well as income from many sources. Families in the survey are interviewed for up to 4 consecutive quarters. 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.[4]

Our main sample is drawn from the quarterly waves spanning the first quarter of 2015 through the second quarter of 2017. Surveys are conducted in each of the three months of the quarter. For each survey, the reference period is the 12 months prior to the interview month for income and the 3 months prior to the interview month for expenditures. Thus, for much of the

[4] A CU is defined as those living together who are related or share resources. Incomes are also only reported in the first and fourth interviews, unless there is a change in the composition or employment status of the CU.

analyses that follow, we will be comparing annual income to four times quarterly expenditures or consumption.[5] For more information on the CE, see BLS (2023).[6]

Administrative Data Sources

We rely on microdata from IRS Form 1040 to obtain adjusted gross income (AGI) for tax filers. The 1040 extracts that we have also contain information on several other income sources, some of which are components of AGI. These include wage and salary amounts, asset income (specifically taxable and tax-exempt interest and taxable dividends), and gross Social Security benefits. Although our 1040 extracts do not contain actual amounts for taxes paid and tax credits received, they cover enough line items (including AGI and filing status) that we can calculate tax liabilities and credits in a relatively accurate manner (Meyer et al. 2022). We also rely on IRS Forms W-2 and 1099-R, which are third-party information returns that contain wage and salary amounts and retirement distributions, respectively. These information returns are primarily used to fill out incomes for individuals who do not file tax returns.

We also use a series of administrative program records that come from various federal and state agencies. To obtain Social Security benefit amounts (Old-Age, Survivor, and Disability Insurance or OASDI), we rely on a combination of universe records from the Master Beneficiary Record (MBR) and Payment History Update System (PHUS) from the Social Security Administration (SSA) (see Logani, Murphy, and Wyse 2023 for further details on the methods for combining these datasets and extracting benefit amounts). To obtain Supplemental Security Income amounts, we rely on universe records from the SSA's Supplemental Security Record file (see Meyer and Wyse 2023 for further details on the methods for cleaning these data). OASDI and SSI are all paid out at the individual level.

We rely on the Public and Indian Housing Information Center (PIC) and Tenant Rental Assistance Certification System (TRACS) files from the Department of Housing and Urban Development (HUD) to obtain estimates of housing assistance (Corinth, Meyer, and Wyse 2023). 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 Agriculture

[5] However, as a robustness check, we will also compare annual income to annual expenditures by constructing annual measures of spending by summing over four quarters of spending for CUs that remain in the survey for all four quarters.

[6] More information on the CE can be found in BLS (2023) and https://stats.bls.gov/cex/.

(USDA) or states and localities. We also use Supplemental Nutrition Assistance Program (SNAP, formerly food stamps) data from state agencies that are available for 22 states during our time period of interest. Housing assistance and SNAP benefits are paid out at the assistance unit level.

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. Linkage error can stem from missed links and wrong links (as pointed out on p. 187 of NASEM 2024), but little is known about the frequency and nature of wrong links.

Because the administrative data for income are typically only available for a given calendar year, we restrict our CE sample to survey months for which the reference period for income closely aligns with the previous calendar year. Specifically, we restrict the sample to include surveys conducted in January through April for survey years 2015 to 2017, which roughly aligns with calendar years 2014 to 2016.[7] The CE includes 77,180 CUs from survey years 2015 to 2017. Restricting to interviews in January through April and to first and fourth interviews (during which incomes are reported) yields a sample of 12,898 CUs. After further restricting our main sample to CUs that include at least one member with a PIK and have an unambiguous state indicator,[8] we obtain a primary analysis sample of approximately 11,200 CUs.

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 member of the CU has a PIK. Data Appendix C provides further details on the inverse probability weighting procedure that we use to match population totals. Since in estimating the weights we use survey reports of our key concepts (income and expenditures), the role for bias from omitted determinants of these outcomes correlated with being PIKed is small. For un-PIKed individuals who remain in our sample (because

[7] Because of the panel structure of the CE, some CUs (about 38%) appear twice in our data. This only impacts standard errors, which are not disclosed.

[8] An ambiguous state indicator is defined the lack of both survey and administrative state data. The state indicator is relevant for the purposes of simulating state tax liabilities and linking administrative SNAP benefits.

they are part of CUs with at least one PIKed member), we continue to use their survey values in cases where we cannot link any administrative data.

3. Methods: Combining Survey & Administrative Data in the Resource Measures

This section discusses how we define and construct our main resource measures. Our main income concept incorporates 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 discuss the methods involved in converting reported expenditures to consumption.

Defining Income

We construct two versions of our post-tax, post-transfer income concept: one using survey information only and another using a combination of survey and administrative inputs (which we call blended income). We start with pre-tax money income, whose components include: 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. Tax liabilities and credits primarily include 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. Our income concept is nearly identical to that of the Congressional Budget Office After-Tax and Transfer Income (see Table 2-1A of NASEM 2024) and is intended to match expenditures (aside from net saving).

Calculating the survey income concept is relatively straightforward, as we simply take the sum of survey-reported amounts corresponding to most income sources. We simulate tax liabilities and credits using the National Bureau of Economic Research's TAXSIM calculator, using inputs from the survey data on state,[9] household structure, incomes, and expenses. We also impute amounts for survey housing assistance (for which only an indicator for receipt is reported) by

[9] In certain instances, the survey omits a state indicator for CUs that have corresponding state indicators in administrative data. In those cases, we take the population-weighted mean of tax liabilities and credits across all states.

subtracting rent paid from an imputed rental equivalent (see Data Appendix A for more details on estimating the rental equivalent). Data Appendix E contains the specific survey variables we use for each income source and additional details on imputing amounts for housing assistance.

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

Constructing Blended Pre-Tax Money Income

We start by describing how we construct a blended version of pre-tax money income. The vast majority of consumer units (CUs) file tax returns. For these CUs, we use Adjusted Gross Income (AGI) reported on IRS Form 1040 as a starting point.[10] AGI comprises 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 itself has shortcomings. It is net of certain deductions and may also miss certain jobs (e.g., if W-2s are left off the 1040 – see Meyer and Wu 2024), informal earnings, and other cash sources. As a result, we rely on the survey-reported analog of AGI if it is both higher and reflects income that is plausibly missed in the administrative records.[11]

A key challenge is that we cannot perfectly align the 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) or non-AGI components being bundled with AGI components as part of a survey question. We therefore construct a modified version of administrative AGI that enables us to conceptually match the union of selected income responses in the survey, and this will serve as a key component of our overall post-tax and post-transfer income concept. We call this concept "modified AGI", which – using the administrative data – is simply AGI plus non-taxable interest (from IRS Form 1040), deferred compensation for wages/salaries (from IRS Form W-2), and

[10] In other work (e.g., Meyer and Wu 2024) using either the Current Population Survey Annual Social and Economic Supplement (CPS ASEC) or the Survey of Income and Program Participation (SIPP), the authors start with the survey income concept and replace survey values with administrative or blended values component-by-component. We use a different approach here – i.e., starting with an aggregate administrative concept like AGI – for at least two reasons. First, several administrative datasets available for the CPS and SIPP are not available for the CE, such as the Social Security Administration's Detailed Earnings Record (DER) which contains administrative 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 at least one of those bundled components are not available in the administrative data.

[11] We also take the sum of wages and deferred compensation reported on IRS Form W-2 and retirement distributions reported on IRS Form 1099-R if this total amount exceeds AGI.

veterans' benefits (from administrative VA data) minus taxable OASDI.[12] We use the survey value of modified AGI when it is higher only if survey earnings amounts are non-imputed and at least one of the following conditions holds:

- Administrative data (from 1040s or W-2s) are missing or zero for the entire CU,
- At least one member of the survey CU indicates that their primary source of income in the past twelve months came from self-employment,
- At least one member of the survey CU indicates that they work in a "high-tip" industry.[13]

In other words, we rely on the higher survey income value only if it reflects income that is plausibly missed in the administrative records (e.g., informal and/or cash-only earnings) and is based on actual, rather than imputed, responses in the survey.

There may be concerns that taking the maximum of two variables with (classical) measurement error would lead one to overstate income. However, it is plausible to assume that the administrative measures are nearly always understated. For example, administrative records may not be linkable if individuals are un-PIKed in the survey. Furthermore, administrative income sources may be either incomplete (e.g., tax records only capture formal sources of earnings) or missing altogether (e.g., in the case of child support, for which we have no administrative equivalent). Even though the survey measures can be either over- or understated, we know on net that they are understated, and we only bring them in when we most strongly suspect the administrative values to be understated. Consequently, we are not particularly concerned about overstating income as a result of these methods. Furthermore, we compare our blended amounts for various income sources to publicly available aggregates in the National Income and Product Accounts (NIPA) as a way of validating these methods ex post. As described in NASEM (2024), it is important to have income measures that can be compared to aggregates in national accounts.

Note that these aforementioned methods pertain to constructing a blended measure of modified AGI for CUs containing only tax filers. For CUs containing only non-filers, we do not

[12] The CE asks about wages and interest inclusive of deferred compensation and non-taxable interest, respectively, and it bundles VA payments in a category that also contains UI, child support, and alimony. Conversely, we want to account for gross OASDI benefits (which we have available in a separate administrative source) but AGI includes only the taxable portion, so we subtract it out to prevent double-counting when we subsequently bring in the gross amount. Backing out the taxable portion of OASDI is an involved process, and more details can be found in Data Appendix D.

[13] We define high-tip industries as sales, retail, private household service, construction, mechanics, and farming. Note that this is likely a conservative correction, since higher survey responses may also be more reliable in other situations (e.g., high-tip occupations in addition to industries, other nontraditional arrangements) and may not be brought in.

have an administrative measure of AGI because there are no 1040s. We therefore begin with the survey measure of modified AGI and make component-by-component substitutions using third-party information returns (specifically wages/salaries from IRS Form W-2 and retirement distributions from IRS Form 1099-R) to construct our blended measure of modified AGI. For the remaining number of CUs that contain a mix of filers and non-filers, we use a combination of the methods outlined above depending on whether the relevant subset of the CU consists of filers or of non-filers.

Finally, after constructing a blended measure of modified AGI for all CUs, we bring in additional money income sources that are not explicitly accounted for in AGI. These include administrative amounts for OASDI and SSI, although we continue to rely on survey values for un-PIKed individuals and when the administrative amounts are imputed.[14] We continue to use survey values for TANF, because the administrative data cover only a subset of states. Data Appendix F provides more details on constructing blended amounts for each of our individual income sources.

Constructing Blended Tax Liabilities and In-Kind Transfers

We now discuss the methods associated with calculating blended measures of tax liabilities and in-kind transfers. We use TAXSIM to simulate taxes using inputs from a combination of IRS tax records and other sources (Feenberg and Coutts 1993). We rely on AGI and tax unit structure (e.g., filing status, number of dependents) from the 1040s and estimate the number of qualifying dependents for a given tax credit by linking birth dates from the SSA's Numerical Identification System (Numident). We calculate payroll taxes on 1) the maximum of wage and salary income from W-2s and the survey (since Social Security taxes are capped at the individual level and payroll taxes are collected even for non-filers) and 2) survey-reported values self-employment income (we lack component level administrative data for self-employment income). We also calculate taxes for families and individuals who do not appear in the 1040s (in the event that they are unlinked, filed late, or had taxes withheld), relying on their survey family structure and on incomes from other IRS information returns or the survey. To determine the state used for calculating state

[14] To obtain a comprehensive administrative record of OASDI benefits received from 2008 to 2018, we combine the 2015 Master Beneficiary Record (MBR) to the 2019 Payment History Update System (PHUS). The former contains indicators for benefit type while the latter contains payment histories from 1984 to 2000 for all beneficiaries. We impute payment amounts for individuals who received benefits as of the MBR extraction date in December 2014 but not as of the PHUS extraction date in October 2019.

income taxes, we use Master Address File (MAF) IDs as well as tax records. The importance of the MAF is further described in NASEM (2024).

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. This may overstate housing assistance if there are false positives in the survey, but housing assistance may still be understated on net if there are false negatives associated with non-HUD programs. For SNAP, we bring in administrative data for 22 states and restrict our analysis to those states when using administrative SNAP values.

Constructing Expenditures and Consumption

The CE collects information on household expenditures for a large number of spending categories.[15] Our measure of expenditures includes all spending collected as part of the Interview survey.[16] To convert reported expenditures in the CE to a measure of consumption, we make a number of adjustments. Our consumption concept is Meyer/Sullivan consumption as summarized in Table 2-1B in NASEM (2024).[17] First, we convert vehicle spending to a service flow equivalent.[18] 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 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. Data Appendix A contains a description of these imputation process. Finally, to arrive

[15] While the CE provides a comprehensive measure of consumption, its spending-based definition may miss non-monetary components, such as free school meals, bartered goods, or household production, which may be relevant for low-income households. This could lead to an understatement of actual consumption, particularly at the bottom of the distribution, highlighting an area for potential improvement through expanded data collection or integration.

[16] Specifically, we define total quarterly expenditures as the sum of the CE variables: TOTEXPPQ and TOTEXPCQ.

[17] As laid out in Table 2-1B of NASEM (2024), our consumption measure differs primarily from the BLS/CE (proposed) consumption measure in that it includes furnishings and durable household equipment but excludes certain education services, in-kind receipts and interhousehold transfers, and home production.

[18] Instead of including the full purchase price of a vehicle or the finance charges on a vehicle loan, we include a flow that reflects the value that a consumer receives from owning a car during the period that is a function of a depreciation rate and the current market value of the vehicle. To determine the current market value of each car owned, we rely on information on the reported purchase price of the vehicle, or we impute the value using detailed information on vehicles (including make, year, age, and other characteristics). See the Data Appendix A for more details on how we calculate vehicle service flows.

at our measure of total consumption, we exclude spending that is better interpreted as an investment such as spending on education and health care, and outlays for retirement including pensions and Social Security.[19] We 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 (more details on our measures of consumption are in the Data Appendix A). Similar arguments have been made by ITWG (2021) and NASEM (2023, 2024), and our preference more broadly is to exclude health insurance from the resource measures that we consider.

4. Results: Comparing Weighted Totals to Public Aggregates

As a precursor to our main results, we compare the weighted totals of blended and survey income components to benchmarks obtained from publicly available aggregates from the National Income and Product Accounts and other sources. This exercise, which follows the approaches used in several previous studies (Meyer, Mok, and Sullivan 2015; Rothbaum 2015; Corinth, Meyer, and Wu 2021), sheds light on the extent of aggregate survey underreporting and provides a way of validating our methodology for blending survey and administrative income. We calculate the weighted total for each income measure using adjusted survey weights (See Data Appendix C). We also make additional adjustments to account for differences between the populations covered by the survey and blended measures and by the aggregate measures.[20]

Figure 1 shows the three-year average reporting rates for key components of the survey and blended income measures when compared against the corresponding aggregate concepts.[21] These results indicate that all our survey-based income measures suffer from underreporting. 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 (averaging 92% for our sample period), the reporting rates vary widely across other components. For example, the rate is high for survey-based OASDI (88%) but low for other income sources such as retirement and pensions (27%), other regular income (31%), and SSI

[19] We also exclude spending on charitable contributions and spending on cash gifts to non-family members. This category is very small relative to total consumption.
[20] For more information on aggregate sources and the adjustments made, see Appendix Table A4.
[21] Yearly reporting rates for other components of income – including AFDC/TANF, SNAP, housing assistance, interest and dividends, rental income and royalties, and self-employment income – can be found in Appendix Table A5, A6a, A6b, A6c.

(49%).[22] In the CPS in 2016, the survey reporting rates are similarly high for earnings and OASDI (each 90%) and higher for sources such as retirement income (55%) and SSI (91%) (Meyer and Wu 2024).[23]

The reporting rates for our blended (CID) income measures are not only higher than those for the survey-only measures 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%. Even for our blended measure, however, some components have low reporting rates. For example, the rate for our blended measure of other regular income is only 56% of the corresponding administrative aggregate.[24] On net, the results in Figure 1 suggest that our blended income measure addresses some of the underreporting that is evident in survey income and therefore provides a more accurate picture of economic well-being. However, our blended income measure still misses administrative records for a number of income sources that are likely to be underreported in survey data, including Unemployment Insurance, child support, TANF, etc.

Because after-tax measures of income should more closely reflect economic well-being (since one spends out of after-tax income), we also consider how well tax components of income compare to administrative aggregates. Figure 2 plots the average reporting rates for our estimates of these tax components using inputs derived from our blended income sources. The reporting rate for these tax components of income are quite high – all but one has a rate above 90%. The one exception is state tax liabilities, which have a reporting rate of 81%.[25]

5. Results: Comparing the 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. Families with few resources typically do little saving and borrowing, so actual expenditures for

[22] Additional survey income components that also have very low average reporting rates include self-employment (57%), interest and dividend income (11%), and cash welfare (14%) (see Appendix Table A5).
[23] Note that we also adjust for payments to those abroad, in institutions, and decedents who are not in the survey frame.
[24] One explanation for this low rate is that while the survey concept of other regular income combines components such as veteran benefits, worker's compensation, child support payments, unemployment insurance, and alimony, we only have administrative data for veterans' disability compensation. Consequently, we must rely on the value of the corresponding survey variable for all other components within this blended measure (see Appendix Table A4 for sources of public aggregates).
[25] One reason as to why our simulated state tax liabilities are particularly low is that we are missing state taxes for observations with an ambiguous state indicator.

those at the bottom should be very similar to actual income – in part justifying our focus on families around poverty.[26] In practice, however, reported expenditures tend to exceed reported income for families at the bottom – likely due to underreporting of income (Meyer and Sullivan, 2003, 2011).[27] By comparing our income measures to expenditures, we can examine the extent to which blended income narrows the gap between income and consumption for families with few resources.

We start by comparing the univariate distributions of survey and blended income to each other and to that of expenditures. We then examine the joint distribution of resource measures before investigating the extent of individual-level discrepancies between income and expenditures. To account for differences in family size and composition, we scale our measures using the SPM three-parameter equivalence scale.[28] Dollar figures are expressed in real 2016 dollars using the Personal Consumption Expenditures Price Index (PCEPI).

Univariate Distribution of Resource Measures

Figure 3 plots the bottom fifty percentiles for three resource measures–blended income, survey income, and total expenditures–pooled across all three years in our sample (also see rows 1, 4, and 7 of Table 1).[29] Consistent with previous research using CE data from earlier years, expenditures exceed survey income at very low percentiles (Meyer and Sullivan 2011).[30] The difference is substantial at very low percentiles. For example, at the 3rd percentile, total expenditures are 140% higher than survey income while only 6% higher than blended income. Given the comparisons to aggregates discussed 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. That expenditures exceed blended income for the bottom 3 percentiles might

[26] This point relies on the household budget constraint; see Chapter 2 (p. 65) of NASEM (2024).
[27] This could also be due to debt repayments or inter-household transfers such as remittances.
[28] A description of this equivalence scale can be found in Data Appendix B.
[29] Our measure of blended income uses administrative data on SNAP only for those living in the subset of states for which we have these data. For those in other states, we rely on SNAP values reported in the CE. A consequence of this is that for the full sample, our blended measure of income includes the value of survey-reported SNAP benefits.
[30] The results in Figure 3 show that expenditures exceed income through the 10th percentile of these distributions, while Meyer and Sullivan (2011) find that expenditures exceed income through the 20th percentile (Table 2). That earlier study, however, examines a sample of "complete income reporters" and relies on CE survey data from years when imputed income values were not available, so missing values for income components were treated as zeros.

suggest that blended income understates resources for families at the very bottom. It is worth noting that our blended income measure still does not account for all resources available for consumption (including in-kind transfers other than SNAP) and may still understate even the components of income that are explicitly included. On the flip side, comparisons of expenditures to national account aggregates indicate that expenditures are underreported in the CE (Bee et al. 2015), so further corrections to blended income may be offset by potential corrections to the underreporting of expenditures.

While our primary focus is on the bottom half of the distribution, it is worth briefly noting patterns in the top half as shown in Appendix Figure A1. Blended income consistently exceeds survey income throughout these higher percentiles, reflecting improvements from incorporating administrative data. At the very top of the distribution, survey income and expenditures converge, but blended income is substantially higher, suggesting that survey income likely underreports certain components such as business and capital income. At the same time, our measure of blended income is likely to be understated at the top as well, since it likely misses some unreported and non-taxed income from capital. While these findings reinforce the value of blended data, a full exploration of top-end discrepancies is beyond the scope of this paper.

Joint Distribution of Resource Measures

One explanation for differences between income and expenditures at specific percentiles is that families at a given point in one distribution are not necessarily the same families at that point in another distribution. To compare income and expenditures for the same families, we also examine the joint distribution of these resource measures, whose importance is highlighted in Conclusion 3-3 of NASEM (2024, p. 124). In Table 1, we report the means of our resource measures for families below specified percentiles of a given resource measure. For example, we report mean resources for families below specified percentiles of expenditures in rows 3, 6, and 8.

Examining 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. This expenditure-income difference in the left tail is consistent with evidence provided in several previous studies (Meyer and Sullivan, 2011; Brewer et al. 2017). We should note that it is not surprising that expenditures exceed income in these comparisons, because we are conditioning on low values of survey income while not restricting to low values of expenditures. When we do the reverse exercise

(conditioning on low expenditures), we see a much smaller difference. For families in the bottom 5% of the expenditure distribution, mean survey income (row 3) exceeds mean expenditures (row 8) by only a factor of 1.75.

The story is quite different when looking at blended income. 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.[31] While nearly two-thirds of the discrepancy is resolved by using blended income, unreported family transfers (Sabelhaus 2024) and savings likely remain as important possible explanations. This pattern suggests that the under-reporting of income can explain much of the difference between expenditures and survey income in the left tail. Given the strong relationship between income and expenditures, especially for those with little borrowing and saving, we would expect expenditures to rise with income. In the results reported in Table 1, however, we do not observe expenditures increasing monotonically with survey income at low levels of income. 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 anomalous result has been found in prior studies (Meyer and Sullivan, 2003, 2011; Brewer et al. 2017). This non-monotonicity is less evident when looking at average spending by *blended* income (row 10). Below the 10th percentile of blended income, expenditures are flat; those below the 3rd percentile of blended income have 1% higher expenditures than those below the 5th percentile, but those below the 5th percentile of blended income have lower expenditures than those below the 10th percentile. That expenditures appear to weakly rise with blended income at very low percentiles suggests that an income measure that incorporates both survey and administrative data is subject to less measurement error. Consequently, for families at the bottom of the income distribution, low levels of blended income appear to be a better indicator of their economic well-being than low levels of survey income.

A similar relationship between income and expenditures is evident in Figure 4, which plots median expenditures in the bottom half of the survey and blended income distribution. Specifically, we show median expenditures by bins of income, where each bin includes approximately 3 percentiles of the income distribution. For example, the leftmost point on the

[31] Gindelsky and Martin (2024) conduct a similar exercise that estimates the joint distribution of disposable personal income (using CPS augmented to NIPA totals) and personal consumption expenditures (from the CE). One difference relative to our approach is that they impute the in-kind value of health care. They estimate that mean expenditures exceed mean income by a factor of 2.14 in the bottom decile of income.

figure reports median expenditures ($25,900) for families in the bottom 3 percentiles of the distribution of survey income. The result in Figure 4 again shows a non-monotonic pattern in expenditures at the bottom of the survey income distribution, notably between the second and third data points (6th and 9th percentiles) and again 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).

Individual-Level Differences in Resource Measures

We also examine the within-family differences between income and expenditures, and the extent to which these differences diminish when using blended income in place of survey income.[32] This exercise further illustrates how improving the measurement of income informs discrepancies between reported income and expenditures, particularly when the latter exceeds the former. Figure 5 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 to 29% when comparing expenditures to blended income. This is consistent with a leftward shift in the distribution of expenditures minus income when replacing survey reports of income with blended values.[33]

We replicate these distributions of expenditures minus blended and survey income across various subsamples of the data, including for those 65 and over (Figure 6a), families having less than $50,000 in survey-reported income (Figure 6b), residents of states with available administrative SNAP data (Figure 6c), and those interviewed in the fourth wave of the CE that have at least one prior interview (Figure 6d).[34] Across each of these different groups, we continue

[32] When conducting any comparison between survey and blended income to expenditures, it is important to consider that the survey-only income is afforded an advantage as expenditures are used as an independent variable within the income imputation procedure.

[33] Interestingly, we also find that the distribution for the difference between expenditure and blended income appears denser at the tails, indicating more frequent occurrences of either large negative or large positive differences between expenditures and blended income. In other words, blending income seems to sharply increase or reduce the income of certain CUs, with the large reductions potentially being a result of our blending procedure failing to capture income which is reported in surveys but missing from administrative data sources. For example, certain income sources such as self-employment income may suffer from incomplete coverage in administrative tax records.

[34] By restricting to states with available administrative SNAP data, we can incorporate administrative SNAP records into our blended income measure and thereby more accurately capture resources for those at the bottom. Additionally, we focus on those who have been interviewed in the fourth wave as they allow us to construct a measure of expenditures based on expenditures over a full year. If the CU in the fourth wave has completed all three preceding interviews, we compute the annual expenditure as the sum of quarterly expenditures across the four interviews. For

to observe fewer individuals whose expenditures exceed income after switching from survey to blended income. Specifically, we observe a 33.2% drop for those 65 or older, a 35.4% drop for those with reported survey income below $50,000, a 22.7% drop for those in states with administrative SNAP data, and a 27.5% drop when we restrict our sample to CUs interviewed in the 4th wave. Using blended income continues to lead to a leftward shift in the distribution of expenditures minus income for each of these samples.[35]

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 the misreporting of income or the presence of saving and dissaving. The large differences between income and expenditures that remain at the bottom of the income distribution (where there is relatively little saving and dissaving) point to the continued persistence of measurement error.

6. Results: Income and Consumption Poverty Measures

Our analyses so far have compared the distributions of income and expenditures, showing that blended income exceeds survey income at every percentile and thus does a better job of matching expenditures at the bottom of the distribution. In this section, we focus on applications to deep poverty and poverty, which allow us to summarize the resource distribution at particular thresholds. Specifically, we calculate the share of individuals in CUs below the poverty line (and half the poverty line) across our measures of survey income, blended income, and consumption. All resource measures are adjusted using the SPM three-parameter equivalence scale and normalized to a family of two adults and two children.[36] While we previously examined expenditures, we focus on consumption in this section 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. To establish our poverty line, we take the weighted

CU one or two incomplete prior interviews, we determine we annualize the sum of expenditures across completed interviews.

[35] We observe an increase in the density in the extreme right tail after replacing survey income with blended income across the majority of our subsamples. The exception is that when we condition on those with survey income less than 50,000 dollars, we fail to see a significant uptick in the density of expenditures minus the blended income measure at the right tail. This suggests that this phenomenon may be a result of our blending procedure assigning very low incomes to CUs with higher survey incomes.

[36] 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.

average of the SPM threshold, which varies by housing tenure, for reference year 2016 to obtain a single value for 2016 (Fox 2017).[37] 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.

Deep Poverty

We begin by examining how the deep poverty rate – the share of those with resources below half the poverty line – differs across our resource measures for the full sample (Figure 7). One advantage of examining the deep poverty threshold is that any discrepancies between income and consumption measures are likely due to measurement error rather than conceptual differences (given that those at the very bottom tend to do little saving and borrowing). Using survey income alone, 5.3% of individuals are classified as deep poor. The rate of deep poverty falls to 1.9% after using blended income, while the deep poverty rate using reported consumption is 1.8%. In other words, 97% of the gap in the survey-reported rates of consumption and income deep poverty can be explained by the under-reporting of income. These results are consistent with the results in the prior section, which suggest that under-reporting of income can explain much of the difference between expenditures and survey income in the left tail. Moreover, the similarity of the deep poverty rates associated with 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.[38]

In addition to calculating deep poverty rates for the full sample, we also estimate deep poverty among selected demographic subgroups. We divide the overall sample into five mutually exclusive family types (single adults with children, single adults without children, multiple adults with children, multiple adults without children, and CUs containing anyone 65+) and into four mutually exclusive age groups (under 18, 18 to 54, 55 to 64, and 65 or older). Our main analyses focus on three subgroups that are particularly policy-relevant (and which are not mutually

[37] For data on the SPM thresholds, see https://www.bls.gov/pir/spmhome.htm#threshold.

[38] It is important to note that the difference between survey income and the other two resource measures is likely overstated given that the survey income concept does not include the value of housing assistance. However, it is unlikely that the inclusion of the value of survey-imputed housing assistance would have an effect so large as to alter the relationship described above. In results not depicted here, we show that the deep poverty rate of blended income without housing assistance is 2.4%. This is 55% lower than the deep poverty rate associated with the corresponding survey concept.

exclusive): individuals in non-elderly single-parent families, individuals under 18 (children), and individuals 65 and older (the elderly).

At a broad level, we continue to see for each subgroup that deep poverty measured using survey income far exceeds the deep poverty rates measured using either blended income or consumption. 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. The remaining differences between blended income deep poverty and consumption deep poverty are thus smaller but non-trivial for each subgroup. Blended income deep poverty is now 17% lower than consumption deep poverty (1.9% compared to 2.3%) for children, while it is still 44% higher for the elderly (1.3% compared to 0.9%) and three times higher for those in single-parent families (6.1% compared to 2%). While the differences are larger 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 (such as SNAP, TANF, WIC, and child support).

Poverty

Moving to higher thresholds, Figure 8 examines the share of individuals in CUs with resources below the poverty line for the full sample and different demographic subgroups. For the full sample, we find that 13% have survey incomes below the poverty line. This rate falls to 7.3% for blended income, a drop of 44%.[39] In contrast, the consumption poverty rate is 17.9%, which exceeds not only that of blended income but also of survey income.[40] We should expect the difference between consumption and income-based poverty rates to grow further up the

[39] This percentage decline is similar to the 41% reduction in poverty that Meyer and Wu (2024) find after using blended data to measure income after taxes, expenses, and in-kind transfers in the CPS ASEC for reference year 2016. Bee et al. (2023), in contrast, find a 9% reduction in poverty after using blended data in the CPS ASEC for reference year 2018 – but they focus only on pre-tax money income and correct for potentially offsetting errors beyond just misreporting.

[40] Prior studies that have compared consumption and income poverty have shown consumption poverty to be lower than income poverty (Meyer and Sullivan 2012a), but this is due to that this prior work anchoring the thresholds in a given year so that the poverty rates across measures are the same at a given point in time in order to focus on difference in trends rather than levels.

distribution, as families with more resources are more likely to save.[41] Some of this difference could also be due to greater under-reporting of expenditures further up the distribution.

We continue to see survey income poverty exceeding blended income poverty for all three demographic subgroups, with the gap being 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 (as was shown in Figure 1). We also see that consumption poverty is substantially higher than blended income poverty for all three subgroups, although the magnitude of this difference varies considerably across groups. For individuals under the age of 18, consumption poverty is substantially higher than both survey income poverty and blended income poverty. In contrast, for single parents and the elderly, consumption poverty hovers slightly below survey income poverty but still remains higher than blended income poverty.

In the results discussed so far, our blended income measure relies on survey-reported rather than administrative SNAP benefits because we only have administrative SNAP records for a limited number of states. Given the high rates of SNAP underreporting that prior studies have documented (e.g., Meyer and Mittag 2019, Meyer, Mittag, and Goerge 2022, Fox, Rothbaum, and Shantz 2022), our blended income poverty rates would be even lower after adjusting for SNAP underreporting. Indeed, after incorporating administrative SNAP records for the subset of states for which we have such data,[42] we find that income poverty falls by 49% (from 12.6% to 6.4%) after switching from using survey-only income to a blended income measure. This decrease is 5 percentage points larger than the 44% decline without including administrative SNAP in our main sample (Appendix Table A8), suggesting that correcting for the underreporting of SNAP in the CE would further reduce income poverty rates in meaningful ways.

[41] Future work could involve examining the distribution and share of savers and dissavers (those with assets and debt) below each multiple of the poverty line for survey and blended income and expenditures. NASEM (2024) emphasizes the importance of examining the budget identity and using income and consumption to determine saving (and eventually wealth).

[42] When examining poverty rates among individuals in states with administrative SNAP data, we prioritize analyzing results at the poverty line instead of half the poverty line (deep poverty) for two key reasons. First, Census disclosure rules do now allow us to disclose blended income deep poverty rates for several important demographic subsamples. Second, survey income deep poverty rates for households with administrative SNAP data are considerably lower than those for the full sample, meaning the inclusion of administrative SNAP data would be affected largely by a change in the sample composition (Appendix Table A8).

Appendix Table A7 shows results comparing poverty rates at higher thresholds (150% and 200% of the poverty line) using survey income, blended income, and consumption. We continue to observe blended income yielding lower poverty rates than survey income, although the differences in percentage terms decline at higher thresholds. Consumption poverty continues to exceed income poverty at higher thresholds.

7. Discussion and Conclusions

We construct the first after-tax and in-kind transfer income and consumption measures of family resources relying on survey values of income and expenditures that have been linked to comprehensive and accurate administrative tax and program data. The linked data enable us to combine, or "blend", survey and administrative values to construct a measure of income that corrects for the misreporting of key survey income sources (such as earnings, OASDI, retirement pensions, and SSI) while simultaneously retaining demographic detail, family structure, and information on income sources unable to be captured in administrative data. We then calculate the poverty rate and the poverty-reducing effects of key income sources across survey-only income, the improved blended income, and consumption measures to examine economic well-being. Specifically, we calculate the poverty-reducing effects of key income sources by recalculating the share of individuals under the poverty line after excluding a given income source, assuming no changes in behavior.

In this paper, we illustrate that once we adjust for income underreporting in the survey and incorporate tax liabilities, credits, and in-kind transfers, the deep poverty rate, defined as 50% of the poverty line, diminishes significantly to 1.9%, which closely aligns with the consumption-based deep poverty rate of 1.8%. This result sharply contrasts with the deep poverty rate associated with survey-reported income, which remains notably higher at an estimated 5.3%. These findings have two key implications: First, survey-reported income measures tend to present an overestimation of poverty, particularly among individuals with the most acute material deprivation. Second, for those at the bottom of the income distribution, using consumption as a metric to gauge economic well-being can serve as a proxy for employing an income measure that has been corrected for measurement error.

It is important to note several caveats regarding our resource measures which may alter our results and that we hope to address in future work. In the future, we aim to incorporate survey-

reported lump-sum income into our measure of survey-only income. Additionally, past research (Bee et al. 2015) has indicated that expenditures as captured by the CE Interview survey, though not as markedly affected as income, still exhibit underreporting. We intend to investigate how scaling expenditures by reporting rates and using administrative PIC/TRACS data for rent paid by individuals residing in public or subsidized housing can enhance the measurement of consumption and expenditure patterns. Finally, for our blended income measure, we hope to incorporate administrative TANF data, improve the accuracy of our blended tax simulations for selected tax components, and refine our blending procedure to address cases where it appears that a key administrative data source is missing for an individual.

When considering the results that compare the distribution of income to expenditures, we plan to produce further results for various subsets of the population, which may include individuals falling under specific thresholds of blended income, those with minimal savings and dissaving (characterized by low assets and debt), and individuals in different age groups. These analyses of subsets of the population may help explain situations where income and expenditures greatly differ, particularly when dissaving or debt-financed consumption contributes to higher levels of expenditures relative to income. In future work, we could explore whether incorporating savings data from the CE, such as responses to savings and asset questions, might help address discrepancies between income and consumption.

We also intend to produce poverty results with anchored poverty rates for income and consumption measures, thereby allowing for greater comparability across different resource measures. Additionally, to assess whether blended income poverty aligns more closely with consumption poverty at the individual level, we will compare the membership of the different groups classified as poor: the consumption-poor, the survey income-poor, and the blended income-poor. Furthermore, to examine how well these definitions identify the most deprived, we will examine several well-being measures for these groups. These measures could include, but are not limited to, mortality rates, long-term income derived from tax forms, as well as survey-reported home characteristics, appliances, and assets.

The findings in this paper underscore the critical value of a robust data linkage infrastructure, as emphasized in NASEM (2024). Such an infrastructure facilitates the integration of survey and administrative data, enabling more precise measurement of economic well-being by reducing misreporting and improving data consistency. As highlighted in the report, investments

in data linkage not only enhance the accuracy of research but also pave the way for innovative methodologies that can inform evidence-based policymaking and advance understanding of inequality and poverty dynamics. Moreover, as NASEM (2024) points out, a comprehensive measure of wealth, or changes in wealth through saving and dissaving, is essential to fully reconcile differences between income and consumption and to better capture household economic dynamics. Future research should prioritize developing and incorporating such measures, following the report's recommendations, to close the circle on understanding the interrelationships between income, consumption, and wealth.

References

Bee, C. Adam, Bruce Meyer, and James Sullivan. 2015. "The Validity of Consumption Data: Are the Consumer Expenditure Interview and Diary Surveys Informative?" In Improving the Measurement of Consumer Expenditures, edited by Christopher Carroll, Thomas Crossley, and John Sabelhaus. Chicago: Univ. Chicago Press.

Bee, Adam, Joshua Mitchell, Nikolas Mittag, Jonathan Rothbaum, Carl Sanders, Lawrence Schmidt and Matthew Unrath. 2023. "National Experimental Well-Being Statistics Version 1." Census Bureau Working Paper.

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

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): October, F24-F49. https://doi.org/10.1111/ecoj.12334

Brummet, Quentin, Denise Flanagan-Doyle, Joshua Mitchell, John Voorheis, Laura Erhard, and Brett McBride. 2018. "What can administrative tax information tell us about income measurement in household surveys? Evidence from the Consumer Expenditure Surveys," Statistical Journal of the IAOS 34: 513-520.

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: 345-50.

Corinth, Kevin, Gillian Meyer, Angela Wyse. 2023. "Measuring Longitudinal HUD Benefit Receipt with Administrative Data". CID Memo

Davies, Paul S. and T. Lynn Fisher. 2009. "Measurement Issues Associated with using Survey Data Matched with Administrative Data from the Social Security Administration." Social Security Bulletin, 69, 1-12.

Dushi, Irena and Brad Trenkamp. 2021. "Improving the Measurement of Retirement Income of the Aged Population." ORES Working Paper No. 116.

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*.

Feenberg, Daniel, and Elisabeth Coutts. "An Introduction to the TAXSIM Model." Journal of Policy Analysis and Management, vol. 12, no. 1, 1993, p. 189, https://doi.org/10.2307/3325474.

Fox, Liana. 2017. "The Supplemental Poverty Measure: 2016". U.S. Census Bureau. Report Number P60-261

Fox, Liana, Jonathan Rothbaum, and Kathryn Shantz. 2022. "Fixing Errors in a SNAP: Addressing SNAP Underreporting to Evaluate Poverty." AEA Papers and Proceedings, 112: 330-34.

Gindelsky, Marina and Robert Martin. 2024. "The Polarization of Personal Saving." BLS Working Paper #575.

Han, Jeehoon, Bruce D, Meyer and James X. Sullivan. 2022. "Who is Poor, How Poverty has Changed, and Why it Matters: Poverty Measurement in the U.S. and its Implications for Policy." University of Chicago Working Paper.

Interagency Technical Working Group (ITWG) Evaluating Alternative Measures of Poverty, Final Report. 2021. https://www.bls.gov/cex/itwg-report.pdf

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 Improving the Measurement of Consumer Expenditures, edited by Christopher Carroll, Thomas Crossley, and John Sabelhaus. Chicago: Univ. 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 Improving the Measurement of Consumer Expenditures, edited by Christopher Carroll, Thomas Crossley, and John Sabelhaus. Chicago: Univ. Chicago Press.

Logani, Ilina, Connor Murphy, and Angela Wyse. 2023 "Measuring Income from OASDI Benefits with Administrative Universe Data". CID Memo

Meyer, Bruce D., Nikolas Mittag, 2017. Misclassification in binary choice models. J. Econometrics 200, 295–311. http://dx.doi.org/10.1016/j.jeconom.2017.06.012

———. 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 (April), 176-204.

Meyer, Bruce, 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 (September), 1605-1643. http://jhr.uwpress.org/content/early/2020/08/06/jhr.58.1.0818-9704R2

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, February, 44(1), pages 52-87.

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

———. 2012b. "Winning the War: Poverty from the Great Society to the Great Recession". Brookings Papers on Economic Activity, Fall 2012, 133-200.

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

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

———. 2024. "Poverty in the United States." University of Chicago 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 Measuring and Understanding the Distribution and Intra and Intergenerational Mobility of Income and Wealth, University of Chicago Press, 2022, 459-498. https://www.nber.org/system/files/chapters/c14441/c14441.pdf

Meyer, Bruce D., Derek Wu, Matthew Stadnicki and Patrick Langetieg. 2023. “The Reporting of Unemployment Insurance and Unemployment in Survey and Administrative Sources.” Working Paper.

Meyer, Gillian. and Angela Wyse. 2023 "Measuring Income from SSI Benefits with Administrative Universe Data". CID Memo

National Academies of Science, Engineering and Medicine (NASEM). 2023. *An Updated Measure of Poverty: Re(Drawing) the Line.* Washington, DC: The National Academies Press. https://doi.org/10.17226/26825

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. https://doi.org/10.17226/27333

Nicholas, Joyce and Michael Wiseman. 2009. “Elderly Poverty and Supplemental Security Income.” Social Security Bulletin 69, pp. 45-73.

———. 2010. “Elderly Poverty and Supplemental Security Income, 2002-2005.” Social Security Bulletin 70, pp. 1-29.

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 2015-01.

Sabelhaus, John. 2024. “Household Wealth Transfers in the United States.” Working Paper.

U.S. Census Bureau. 2016. “Preliminary Research for Replacing or Supplementing the Income Question on the American Community Survey with Administrative Records,” Memorandum Series #ACS16-RER-6.

Warshawsky, Mark J. 2022. “Social Security reform, part II: Private retirement accounts and pensions.” Blog Post, American Enterprise Institute. March 25.

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.” U.S. Census Bureau.

Tables and Figures

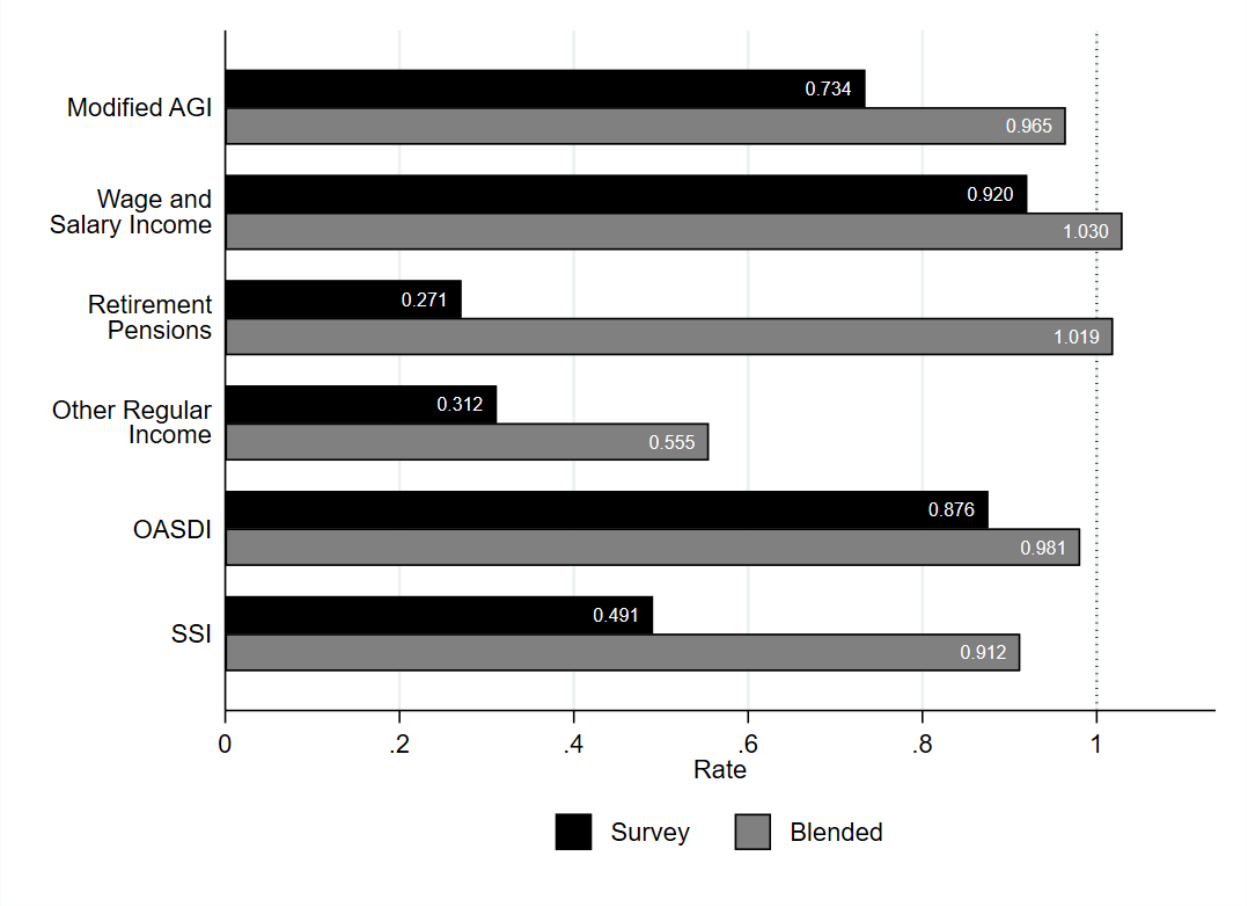
Table 1. Percentiles and Conditional Means of Resource Measures

		Percentiles								
		3rd	5th	10th	20th	30th	40th	50th	75th	90th
		Survey Income								
(1)	Percentile of survey income	6448	12610	22250	32150	40660	49720	59620	93550	143600
(2)	Mean for those below given percentile of survey income	2501	5287	11610	19520	25260	30380	35270	48580	59800
(3)	Mean for those below given percentile of expenditures	21940	24110	27200	33140	36870	41050	45200	57120	66080
		Blended Income								
(4)	Percentile of blended income	14860	20790	28680	40370	50940	63420	75980	119200	188200
(5)	Mean for those below given percentile of blended income	8282	12390	18830	26710	33070	39120	45240	62000	76620
(6)	Mean for those below given percentile of expenditures	33020	37170	39880	46600	50700	55580	60210	74540	86420
		Expenditures								
(7)	Percentile of expenditures	15710	17820	22400	29420	35730	42360	49430	75990	119800
(8)	Mean for those below given percentile of expenditures	11840	13780	17100	21640	25300	28790	32240	42070	51050
(9)	Mean for those below given percentile of survey income	34270	36950	34590	35560	37240	40140	43080	50230	57530
(10)	Mean for those below given percentile of blended income	31340	31060	31440	33480	36650	39660	42630	50430	57520

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 C and the sources used to construct publicly available aggregates are listed in Appendix Table A4. Approved for release by the Census Bureau's Disclosure Review Board, authorization numbers CBDRB-FY23-0184 and CBDRB-FY2023-CES005-016.

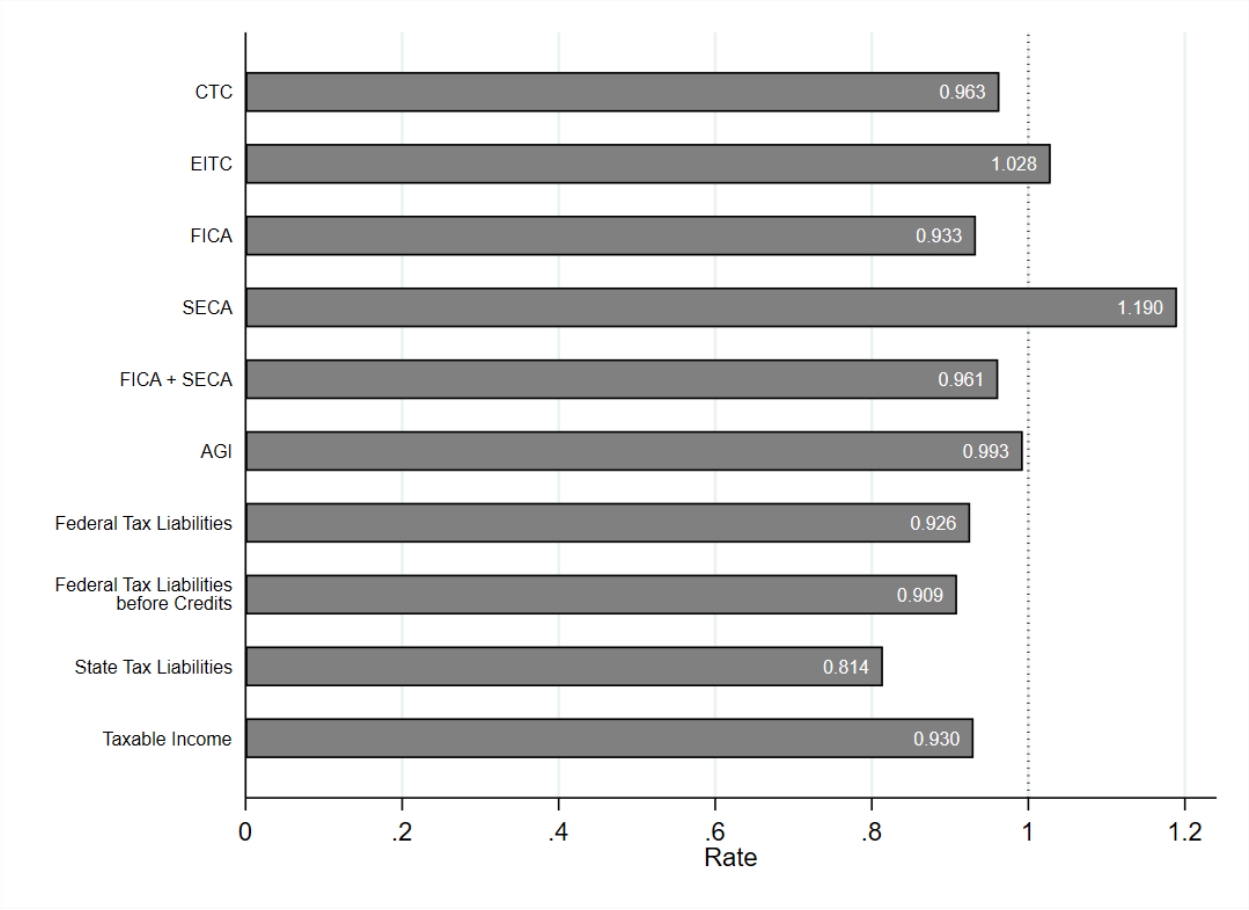
Figure 1. Average Reporting Rates of Blended vs. Survey Income Categories (2014-2016)



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 C and the sources used to construct publicly available aggregates are listed in Appendix Table A4. Modified AGI can be 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.

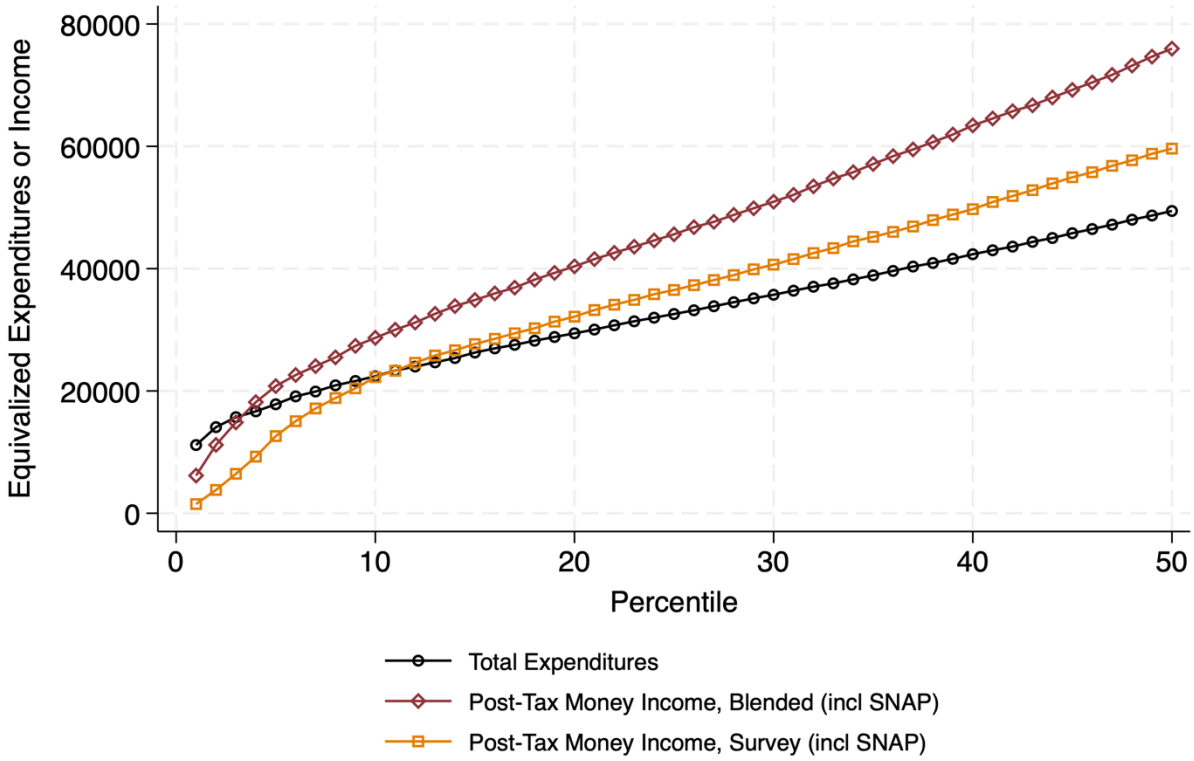
Figure 2. Average Reporting Rates of Tax Estimates Created using Blended Inputs



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 C and the sources used to construct publicly available aggregates are listed in Appendix Table A4. Approved for release by the Census Bureau's Disclosure Review Board, authorization numbers CBDRB-FY23-0184 and CBDRB-FY2023-CES005-016.

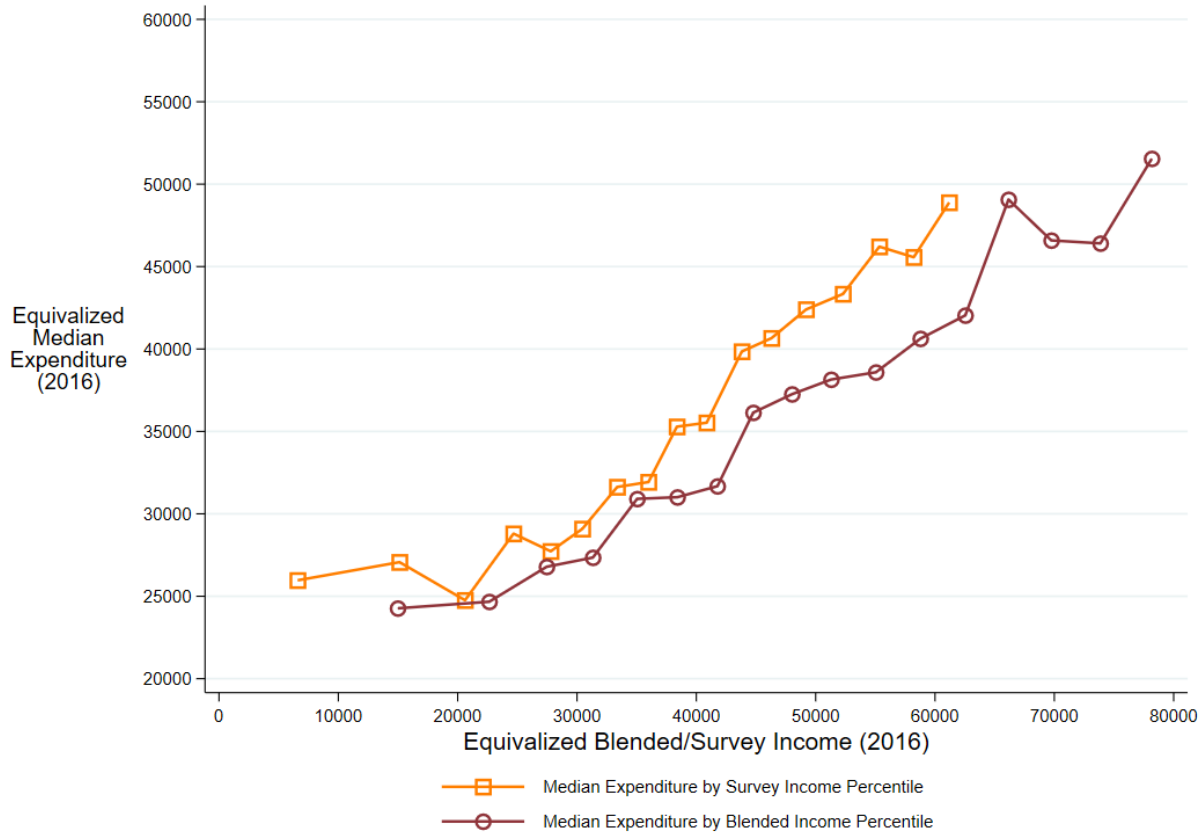
Figure 3. Distribution of Resource Measures, Bottom 50 Percentiles



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

Notes: 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. Approved for release by the Census Bureau's Disclosure Review Board, authorization numbers CBDRB-FY23-0184 and CBDRB-FY2023-CES005-016.

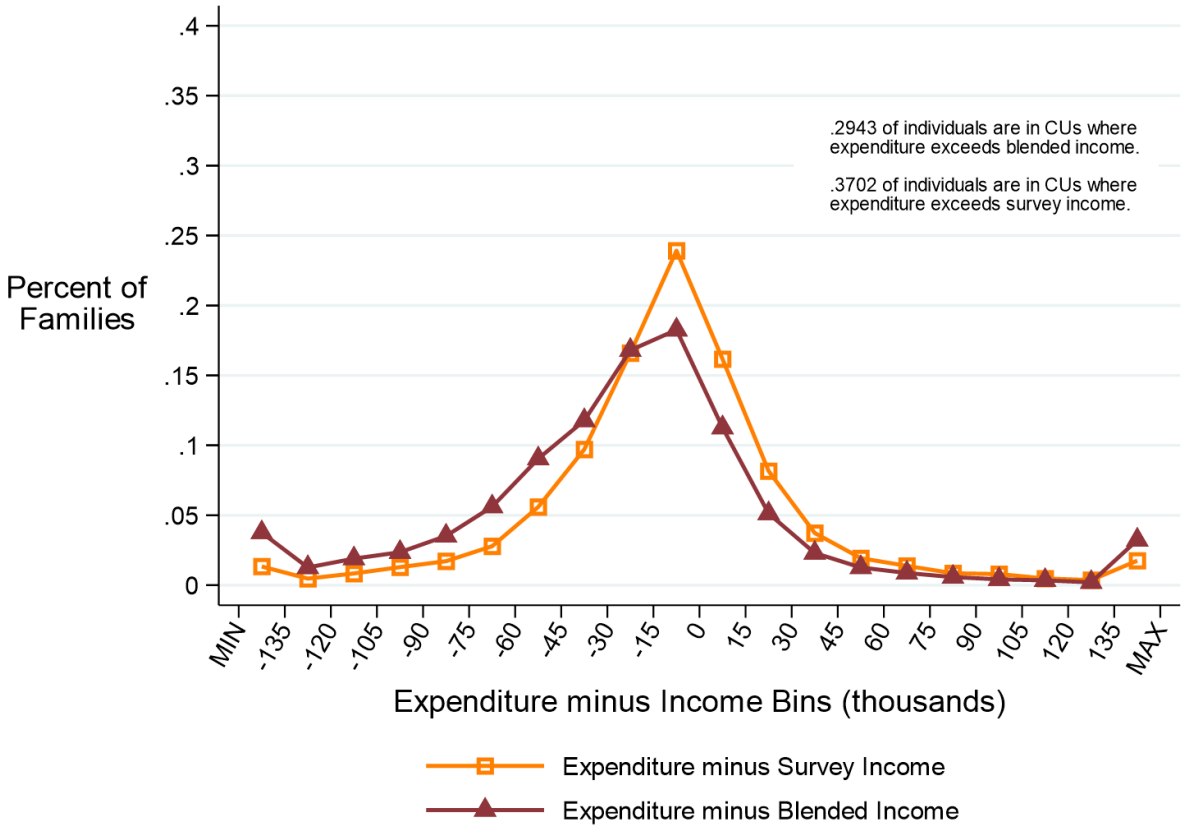
Figure 4. Median Expenditure by Income Quantile



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

Notes: We divide the sample into 33 equally populated bins and display the bottom 17 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 SNAP. Approved for release by the Census Bureau's Disclosure Review Board, authorization numbers CBDRB-FY23-0184 and CBDRB-FY2023-CES005-016.

Figure 5. Distribution of Expenditure minus Income, Full Sample.

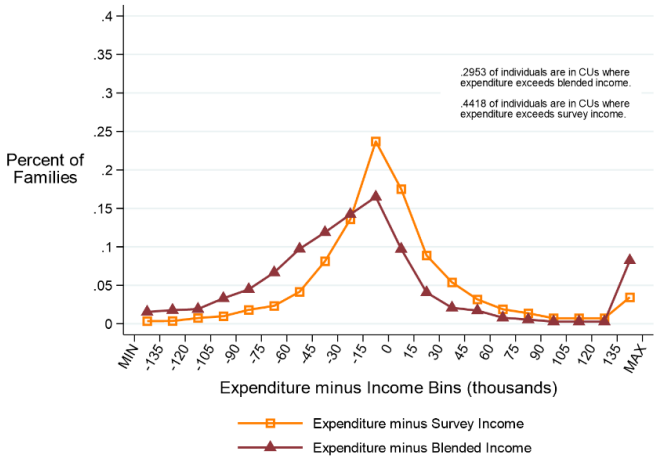


Source: 2014-16 Consumer Expenditure Survey linked to various administrative records

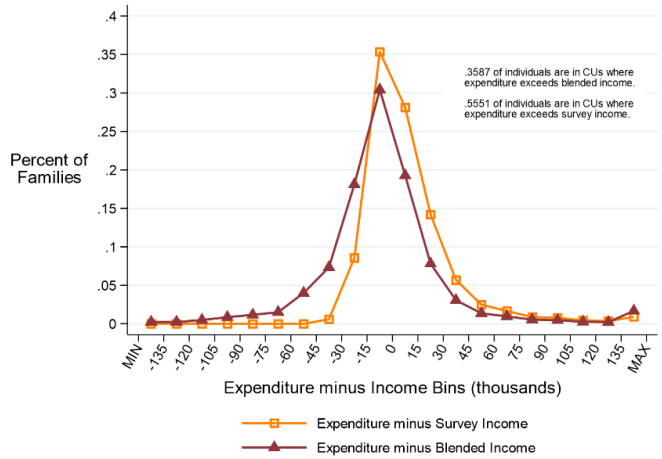
Notes: We partition the data into $15,000 bins and position markers at the midpoints of these bins. All values are expressed in 2016 dollars and adjusted for differences in family size using the three-parameter SPM equivalence scale and normalized to a two-adult and two-child family. Income is after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

Figure 6. Distribution of Expenditure minus Income, Subsamples

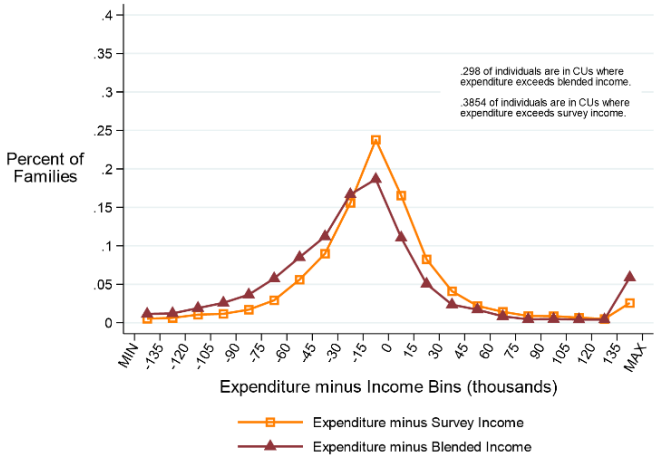
a) Elderly Individuals (65+)



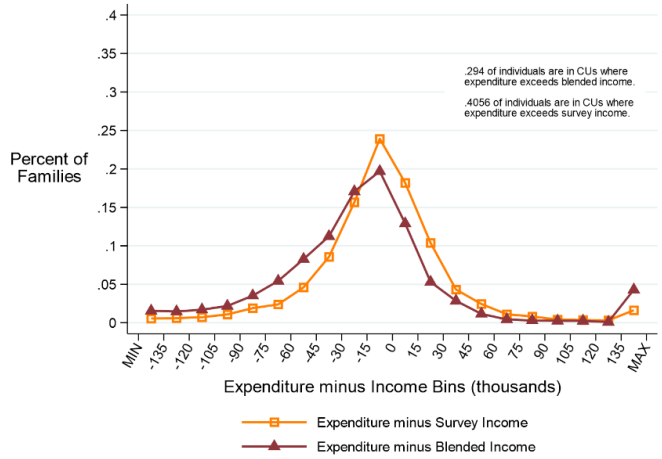
b) Individuals in CUs with Survey Income less then $50,000



c) Individuals in States with Administrative SNAP Data



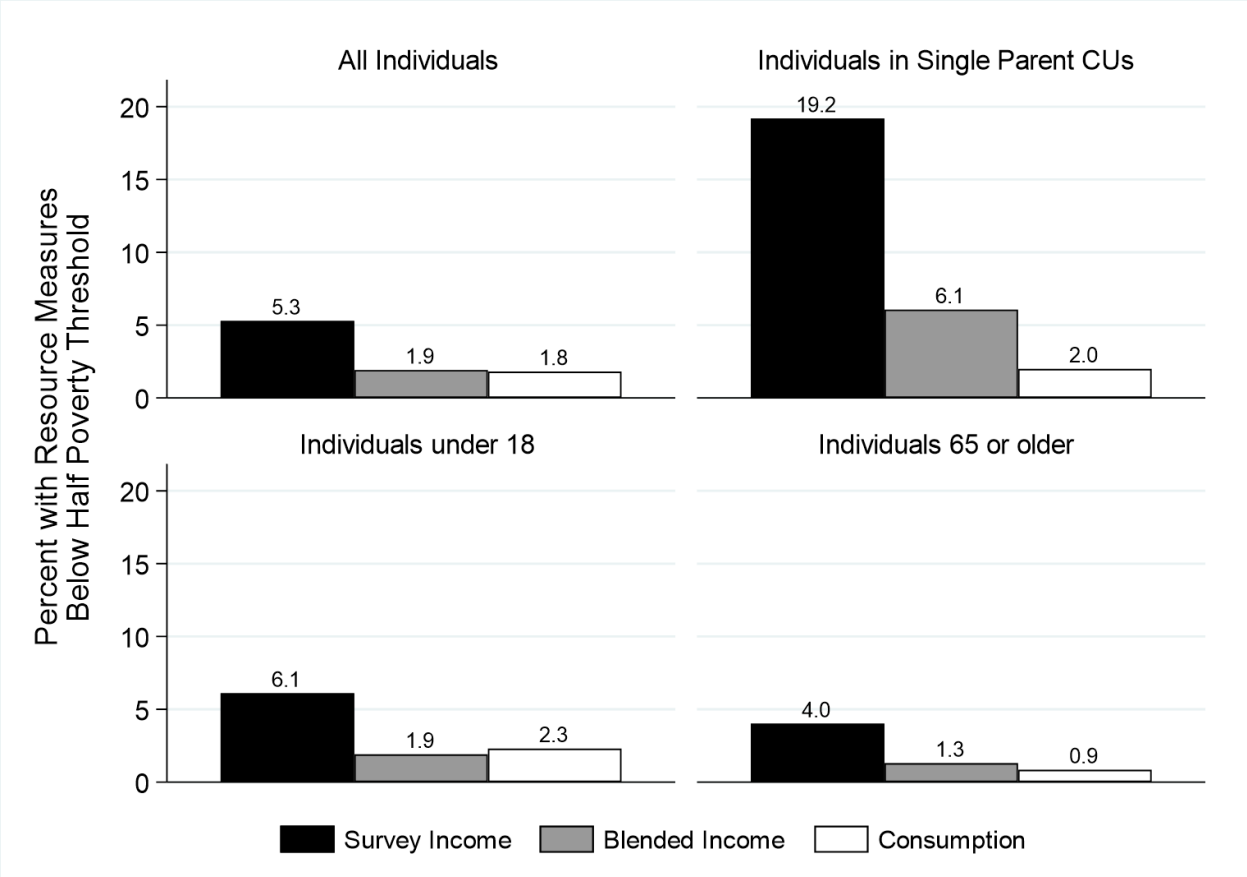
d) Individuals in CUs interviewed in the 4th Wave.



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 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 SNAP. Approved for release by the Census Bureau's Disclosure Review Board, authorization numbers CBDRB-FY23-0184 and CBDRB-FY2023-CES005-016.

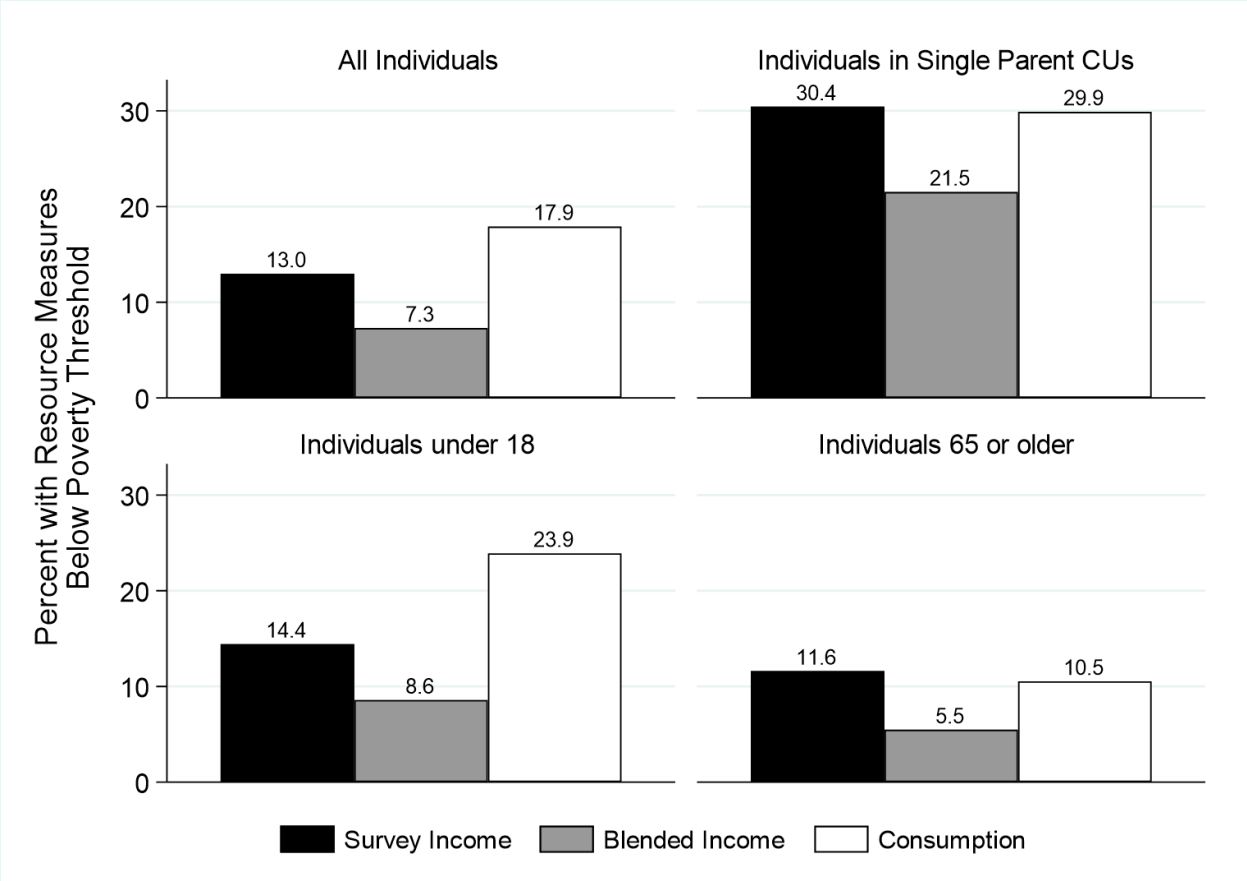
Figure 7. Share of People in CUs below Half Poverty Line, Full Sample.



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

Notes: The survey income concept is post-tax and includes SNAP. The blended income concept is post-tax and includes 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.

Figure 8. Share of People in CUs below Poverty Line, Full Sample



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

Notes: The survey and blended income concepts are post-tax and include housing assistance and SNAP. The blended income measure does not include administrative SNAP benefits, as the data are not available for the full sample of states. 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

	Consumer Unit-Level			Individual-Level		
Reference Year	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. Average Values of Select Demographic Characteristics across Samples

Demographic Characteristic	1st and 4th Interviews, Full Sample (1)	1st and 4th Interviews, SNAP State Subsample (2)	4th Interviews, Full Sample (3)	4th Interviews, SNAP State Subsample (4)
Age of Ref. Person	51.14	51.9	51.53	52.46
Race of Ref. Person	1.361	1.369	1.365	1.365
Education of Ref. Person	13.31	13.41	13.32	13.41
Urbanicity	1.057	1.044	1.058	1.04
Housing Tenure of CU	2.404	2.441	2.373	2.408
Census Division	5.159	4.22	5.132	4.221
Ethnicity of Ref. Person	0.1366	0.1311	0.1335	0.135
Indicator for public housing	1.917	1.904	1.925	1.908
Indicator for subsidized housing	1.936	1.924	1.923	1.92
Marital status of Ref. Person	2.332	2.351	2.309	2.341
Occupation of Ref. Person	6.086	6.055	5.978	5.957
Sex of reference person	1.527	1.519	1.519	1.517

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

Notes: This table presents the weighted mean values of demographic characteristics across the full sample of CUs interviewed in the 1st and 4th wave, the sample of CUs interviewed in the 1st and 4th wave residing in states with administrative SNAP data, the full sample of CUs interviewed in just the 4th wave, and the sample of CUs residing in SNAP states interviewed in just the 4th wave. The weights are described in Data Appendix C. 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 A4. 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) NIPA (Proprietors' income, farm)
Interest and Dividends	Yes	NIPA (Personal interest income) NIPA (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) (Unemployment Insurance) Office of Child Support Enforcement (Child Support) 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 A5. Comparisons of Weighted Survey Income Values to Publicly Available 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 11200 CUs interviewed in the 1st and 4th quarter. The weights are calculated using the methodology described in Data Appendix C.

Table A6a. Comparisons of Blended Weighted Income Values to Publicly Available Aggregates (millions)

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 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 A6b. Comparisons of Blended Weighted Income Values to Publicly Available Aggregates (millions)

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 A6c. Comparisons of Blended Weighted Tax Simulations to Publicly Available Aggregates (millions)

Income Source	CE Blended (1)	Aggregate (2)	Reporting Rate (3)
Panel C. TAXSIM27 Tax Simulations			
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. The likely source of the difference comes from the fact that AGI and taxable income are relatively close to their respective benchmarks suggests that errors in estimating deductions and exemptions, given that AGI and taxable income are relatively close to respective benchmarks are the source of the difference.
Notes: This table presents the weighted totals and reporting rates of blended income components for the sample of 11200 CUs interviewed in the 1st and 4th quarter. The weights are calculated using the methodology described in Data Appendix C. 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 A7. 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 for at deep poverty and near poverty. For other multiples of the poverty line, housing benefits are included.

Notes: This table presents the poverty rates for our main sample of 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. 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.

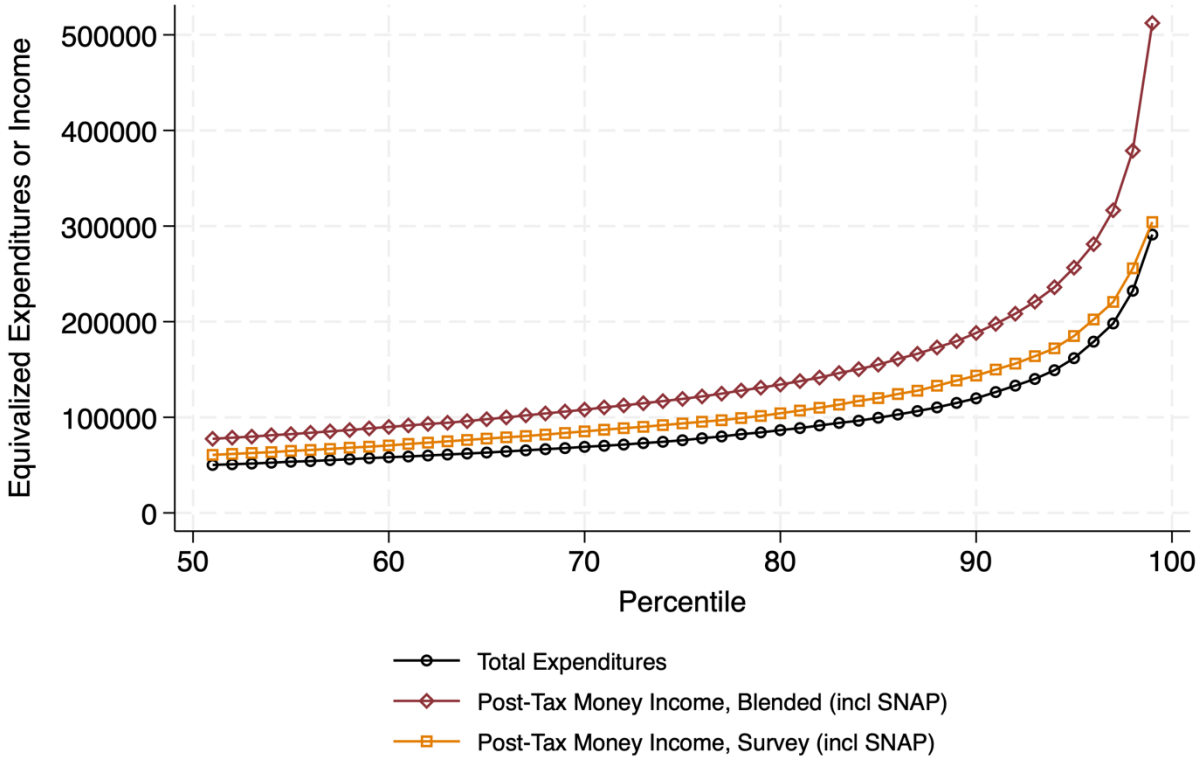
Table A8. Poverty Rates for those in States with Administrative SNAP Data

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 for at deep poverty and near poverty. For other multiples of the poverty line, housing benefits are included.

Notes: This table displays the poverty rates for the sample described in Table A7 (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. Distribution of Resource Measures, 51st-99th Percentiles



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

Notes: 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. 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 – 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.[43]

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

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

log real purchase price for vehicle i, $(x_i\hat{\beta})$. The predicted current market value for each vehicle 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 B – 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.

Appendix C – 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.[44] 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.

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

Appendix D – 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:

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

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[45]. 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.

[45] 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 E - 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[46]	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)

[46] We only exclude the value of Housing Assistance in income when comparing income to expenditures.

Appendix F – Methodology for Constructing Blended Income

Our goal of creating 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 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 filer CUs, the Consumer Unit 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:

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

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.

OASDI

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):

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

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.

SSI

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)

Since we do not yet have cleaned administrative TANF data yet, 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, survey SNAP amounts are utilized. When conducting analyses on the subsample comprising 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.

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

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.

   Salary and Wages

   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 to blended modified AGI 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 procedures for filer CUs.

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 the one outlined in step 1 of the blending procedure for filer CUs.

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

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.

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

The blending methodologies are the same as outlined in step 6 of the blending procedure for filer CUs.