

## **Is the Consumer Expenditure Survey Representative by Income?**

John Sabelhaus<sup>a</sup>  
David Johnson<sup>b</sup>  
Stephen Ash<sup>b</sup>  
David Swanson<sup>c</sup>  
Thesia Garner<sup>c</sup>  
John Greenlees<sup>c</sup>  
Steve Henderson<sup>c</sup>

November 18, 2011

### **Abstract**

Aggregate underreporting of household spending in the Consumer Expenditure Survey (CE) can result from two fundamental types of measurement errors: missing spending for the very highest-income households who are under-represented in the final sample, and underreporting of spending by at least some households who do respond to the survey. Using a new data set linking CE units to zip-code level average Adjusted Gross Income (AGI), we show that the very highest-income households are less likely to respond to the survey when they are sampled, but unit non-response rates are not associated with income over most of the income distribution. Although increasing representation at the high end of the income distribution could in principle significantly raise aggregate CE spending, the low reported average propensity to spend for higher-income respondent households could account for at least as much of the aggregate shortfall in total spending.

<sup>a</sup>Corresponding author, Board of Governors of the Federal Reserve System. Email: john.sabelhaus@frb.gov. <sup>b</sup>U.S. Bureau of the Census. <sup>c</sup>U.S. Bureau of Labor Statistics. Views here do not represent those of the Federal Reserve Board, the U.S. Bureau of the Census, or the U.S. Bureau of Labor Statistics. This paper was prepared for the December, 2011 CRIW/NBER Conference, *Improving the Measurement of Consumer Expenditures*.

## 1. Introduction

Aggregate spending in the Consumer Expenditure Survey (CE) is well below comparable Personal Consumption Expenditures (PCE) in the National Income and Product Accounts (NIPA), and the ratio of these has fallen from where it was two decades ago.<sup>1</sup> Assuming NIPA values are a good benchmark, two potential reasons for the aggregate spending difference are that higher-income families (who presumably spend more than average) are under-represented in the CE estimation sample, or there is systematic under-reporting of spending by at least some CE survey respondents.<sup>2</sup> Resolving why the aggregate shortfall occurs is important for weighting the Consumer Price Index (CPI) and for various research questions that involve the joint distribution of spending and income, including measuring inequality, studying savings behavior, and evaluating the distributional burden of consumption taxes.

Establishing the basic facts about the accuracy of aggregate CE spending is straightforward in principle, but complicated in practice because the CE and PCE differ in terms of both spending concepts and population coverage.<sup>3</sup> However, piecing together the latest Bureau of Labor Statistics (BLS) estimates with the results of a study by Garner, McClelland, and Passero (2009) provides a compelling story (Table 1). There are systematic differences across types of spending at any point in time, and there is also a general decline in the ratio of CE to PCE by about 10 percentage points between 1992 and the early 2000s. However, since 2003, the CE-to-PCE ratio has been relatively stable, both overall and within broad categories of spending.

On net, the CE now appears to be capturing 78 percent of comparable PCE, though that overall ratio is being pushed up because the estimated value of owned housing services in the CE

---

<sup>1</sup> Crossley (2009) shows that the same basic conclusion holds for the British equivalent of the CE survey.

<sup>2</sup> As discussed in Garner et al. (2006), there are possible components for which PCE may be overstated.

<sup>3</sup> For example, PCE includes consumption spending by non-profit institutions.

is much higher than in the PCE.<sup>4</sup> The ratios for durables, non-durables, and non-housing services are 60, 64, and 72 percent respectively, so it seems more descriptive to say that, except for housing services, the CE estimates of comparable spending are generally about one-third lower than the PCE. Again, the within-category declines between 1992 and 2003 are roughly proportional, suggesting that the overall spending ratio decline is not attributable to decreased reporting of any particular type of spending.

Although CE is fundamentally designed to collect expenditure data, not income data, a failure to reflect the income distribution accurately could be a symptom of flaws in the spending distribution as well. There is evidence that the CE does not capture as much income as in other surveys, and the missing income seems to be at the top of the income distribution. Passero (2009) shows that the CE aggregate income is only 94 percent of Current Population Survey (CPS) aggregate income.<sup>5</sup> Evidence that the missing CE income occurs at the very highest income levels comes from comparing CE against other data sets. A comparison of the CE income distribution to the CPS, Survey of Consumer Finances (SCF), and tax return-based Statistics of Income (SOI) data sets suggests significant under-representation of the \$100,000 or more income group in the CE. The CE finds fewer households in that income range, and the average incomes for households that are above \$100,000 are well below the averages in the other data sets.

It may be that higher-income CE households are simply less likely to accurately report their actual incomes, but there are good reasons to suspect that the very top of the income distribution is under-represented in the CE. The first type of evidence comes from a new approach to this question developed for this paper. The approach involves linking all CE

---

<sup>4</sup> For a discussion of how owned housing services are estimated in the CE, see Garner and Short (2009).

<sup>5</sup> See also Meyer, Mok and Sullivan (2009).

sampled households (both respondents and non-respondents) to the average Adjusted Gross Income (AGI) in their five-digit zip-code area. For most of the AGI distribution there is little or no association between unit non-response and zip-code level AGI, but at the very top of the income distribution the unit response rate and the ratio of average CE income to mean zip code-level AGI are both lower. That is, in the top few percentiles of households sorted by zip-code level AGI, households are less likely to participate in the CE, and those households that do participate are more likely to have incomes below the average in their zip-code.

While the CE seems to be missing households at the very top end of the income distribution, under-reporting of spending for at least some respondents is also quite likely. This argument is based on the observation that BLS-reported total expenditures are already lower than total after-tax income, and although some of that cash-flow residual is actual saving and some is attributable to known measurement errors in CE income taxes, those explanations alone cannot reconcile the cash-flow residual. Moreover, the ratio of spending to after-tax income falls with after-tax income, and thus the unaccounted-for cash flow seems to be concentrated at the top of the income distribution. Comparison of CE incomes to other data sources suggests a significant share of income is missing in the \$100,000 or more income group, which means the unaccounted-for cash-flow for that group would be even larger if high-income households are well-represented but simply under-reporting income.

Why is it important to distinguish between the possible explanations for under-reporting of aggregate CE spending? The difference in CE to PCE aggregates across the broad categories in Table 1 highlights one key reason—weighting the CPI. If there are systematic differences in how well the CE survey captures aggregate expenditures across categories, the CPI weights will be biased, and the overall index will be inappropriately affected by changes in the prices of over-

or under-represented categories.<sup>6</sup> Given the plutocratic nature of the CPI, the relationship of income and spending on different types of categories suggests that under-representation of high-income families in the CE could be biasing the CPI.

In addition to weighting the CPI, however, there are also several research areas where the ratio of expenditures to income across income groups is the crucial input, and thus discerning between under representation of high-income families versus under-reported expenditures for at least some respondents is crucial. CE data have been used in several studies to measure differences between consumption-expenditure and income inequality, with consumption-expenditure inequality shown to be consistently and dramatically lower.<sup>7</sup> Bosworth, Burtless, and Sabelhaus (1991) used CE data to track changes in household saving across groups and time, and the estimated patterns of low-income dissaving and high-income saving are dramatic in every period. Finally, CE data are regularly used by government agencies and other groups to measure the distributional burden of consumption taxes. Consumption taxes appear very regressive, because the ratio of spending to income falls dramatically with income.

If the source of the aggregate CE shortfall is simply under-representation of the highest-income households, then the inequality, saving, and tax distribution studies described above may be incomplete, but they are not necessarily biased for the range of the income distribution they represent. Even though the very highest-income households are under-represented in the CE, Sabelhaus and Groen (2000) demonstrate that the overall under-reporting of spending is partially attributable to under-reporting of expenditures by at least some CE respondents.<sup>8</sup> If expenditure

---

<sup>6</sup> See McCully, Moyer, and Stewart (2007). See also Blair (2011), prepared for this conference.

<sup>7</sup> See, for example, Johnson and Shipp (1997), Short, et al (1998), Attanasio, et al (2002), Krueger and Perri (2006), Krueger et al (2010), and Heathcote et al (2010).

<sup>8</sup> Sabelhaus and Groen (2000) use a variety of techniques, including appealing to consumption-smoothing theory, to argue that the ratio of consumption to income for high income families is biased down.

under-reporting is indeed worse for higher-income households, then the results of the CE-based inequality, saving, and tax-distribution research should be revisited.

The CE program at BLS is currently in the midst of a multi-year redesign project, called Gemini. The mission of the Gemini project is to redesign the CE in order to improve data quality through a verifiable reduction in measurement error, with a particular focus on under-reporting.<sup>9</sup> Moreover, the National Research Council, through its Committee on National Statistics (CNSTAT), has convened an Expert Panel to contribute to that planned redesign.<sup>10</sup> It is hoped that the results presented in this paper will constitute a further contribution to the CE improvement program.

## **2. How Does the CE Income Distribution Compare to Other Data Sources?**

Although the Consumer Expenditure Survey (CE) estimation sample reflects the actual distribution of households by income over most ranges, comparisons between the CE and other household surveys suggest that the very highest income families are under-represented. In this section weighted counts of CE units and average incomes are compared against three other data sources, the Current Population Survey (CPS), the Survey of Consumer Finances (SCF), and the IRS tax-return based Statistics of Income (SOI). The comparisons include one data set (CPS) that is similar to the CE in sampling strategy, but more focused on income, one that is purely administrative (SOI), and one that employs differential sampling for high-wealth households in order to capture the top of the wealth distribution (SCF).

The overall count of sampled units in the CE, CPS, and SCF are similar. Although the CE samples “consumer units” and the CPS samples “households” and the SCF samples “primary

---

<sup>9</sup> For a description of the Gemini project see <http://www.bls.gov/cex/geminiproject.htm>.

<sup>10</sup> See <http://www8.nationalacademies.org/cp/projectview.aspx?key=49322>.

economic units,” the overall counts for any given year are within 2 or 3 percent (Table 2, last column). The count of units for the SOI is very different from the other surveys, because dependent filers—usually children living in their parents’ home—have to file their own tax returns. There are also differences in the income concept in the SOI, because non-taxable forms of income (mostly transfers) are not included in adjusted gross income (AGI). After adjusting for those differences, though, the four data sets are broadly consistent across the income categories.<sup>11</sup>

The well-known skewness of the U.S. income distribution shows up clearly in the CE as one moves from less than \$50,000 of income (65.1 million consumer units), to between \$50,000 and \$100,000 (34.9 million), to \$100,000 or more (18.9 million). The counts of units for the CPS, SCF, and SOI are shown as differences from the CE values, and the general impression one gets is that the differences are second order. All three data sets show the same basic shape. The SOI, as expected, finds many more units in the less than \$50,000 group, because of dependent filers and the fact that non-taxable transfers are not being included.

The focus of the analysis here is the top of the income distribution, however, and although the counts of units are broadly similar in the \$100,000 or more income category, the total income received by that group is much lower in the CE than in the other three data sets. For example, the CPS finds 22.1 percent more income for those households. Although much of that is because the CPS finds more households above the \$100,000 line, there is no reason to expect

---

<sup>11</sup> Not accounted for in the Table 2 analysis is the fact that CE income includes the value of food stamps and food and rent as pay. Additionally, some people filed tax returns in 2008 because that was a prerequisite to receive the 2008 tax rebate.

any divergence at all between the CE and CPS, because the sampling approach and income concepts (with the exceptions of food stamps income and rent as pay) are similar.<sup>12</sup>

The more noticeable differences in top incomes occur when one compares CE (and CPS) to the SCF and the SOI. The SCF uses an income concept that generally matches the CE, but employs a different sampling strategy in order to capture the top of the wealth distribution.<sup>13</sup> The SCF finds nearly 60 percent more income in the \$100,000 or more income range. To put those numbers in perspective, the nearly \$2 trillion of additional income that the SCF finds at the very top is similar in magnitude to the aggregate spending mismatch that motivates this study.

The conceptual differences between CE and the SOI make direct comparisons more problematic. Using an AGI income concept with the CE data will yield an even lower estimate of income. However, the SOI still finds over 30 percent more income in the \$100,000 or more range even though there are fewer tax filers in that AGI range because of the differences between AGI and the more generalized income concept used in the other surveys. Thus, on net, comparing the CE to both the SOI and SCF data suggests that the very highest income households are under-represented in the CE (and in the CPS, though to a lesser extent).

---

<sup>12</sup> Of course, the one major difference is that the CPS is focused on collecting income, while the CE is focused on spending, and thus income in the CE is an auxiliary demographic control just like age or homeownership. There are also differences between the CPS and CE in terms of imputation and top-coding procedures, though the latter should not be an issue because this comparison is based on published data. See Passero (2009) and Paulin and Ferraro (1996) for a discussion of income imputation in the CE, and Burkhauser et al (2009) for a discussion about how using the CPS without top codes affects estimates of the incomes at the very top of the income distribution.

<sup>13</sup> The calculations in Table 2 are based on internal SCF data, and the income concept excludes capital gains in order to be comparable with the CE/CPS measures. For a general discussion of the SCF see Bucks, et al (2009), and for a general discussion of SCF design and implementation, see Kennickell and Woodburn (1999). The SCF sampling strategy is focused on wealth measurement, but Kennickell (2009) describes how wealth and income are related.



### **3. Why Does the CE Under-Represent the Very Highest Income Households?**

The Consumer Expenditure Survey (CE) is designed to collect expenditure data and related demographic characteristics from a sample that is representative of the U.S. civilian non-institutional population. Currently, the procedures to ensure this representativeness do not account for income. However, if the variables used to produce representative expenditure estimates are highly correlated with income, then the CE random sampling approach should still generate an unbiased representation of the true population income distribution. However, two problems associated with sampling could lead to the under-representation of very high income households.

The first potential problem is sampling variability, because income is highly concentrated at (and even within) the top percentiles, as indicated in both tax data and targeted surveys like the Survey of Consumer Finances (SCF). Sampling variability implies that the estimated aggregates will be very dependent on whether those probabilistically-rare households are chosen to participate in the survey. The fact that CE incomes are systematically lower at the top end—and not just extremely volatile at the top end—implies that sampling variability is not the problem.

The second possible problem is differential unit non-response.<sup>14</sup> The concern here is that the highest income households are less likely to participate in the survey when they are selected. The fact that incomes are systematically lower at the top end of the income distribution in the CE suggests that differential unit non-response among very high income households is an explanation worth exploring, and that is the focus of this section.

There is no direct way to assess whether or not the very highest income families are less likely to participate in the CE when they are chosen, because we do not observe the actual

---

<sup>14</sup> The discussion here follows a long literature on unit non-response. See, for example, Groves (2006) and King, et al (2009) for useful introductions to that literature.

incomes of non-participants. However, it is possible to make indirect inferences about survey participation using a new data set that links sampled CE units to the average Adjusted Gross Income (AGI) in their five digit zip-code area. The average AGI values linked to sampled CE units are produced by the IRS Statistics of Income Division, and are available for public use.<sup>15</sup>

The data set built for this analysis starts with all consumer units selected for the CE for calendar years 2007 and 2008.<sup>16</sup> There are 104,830 units selected for participation, and 74 percent of those participated in the survey. However, the BLS excludes the first (or “benchmark”) interview when publishing expenditure estimates for publication, and that approach is followed here. Thus, the final data set includes 61,546 interviewed respondents out of 83,366 in-scope sampled units, which is an overall response rate of 74 percent.

The analysis here is based on sorting the sampled CE households into income groups using the average AGI for their zip code. This makes it possible to sort both respondents and non-respondents using the same income measure, and to test for differences in response rates across AGI percentiles. Basically, the first step uses the average response rates for the CE sample in each of the 100 AGI percentile-income groups. The second step is to compare the average incomes of respondents to the average AGI for their zip code, again, by AGI percentile. Note that in both steps the percentile-cell calculations all involve several hundred observations

---

<sup>15</sup> See <http://www.irs.gov/taxstats/indtaxstats/article/0,,id=96947,00.html>. Although the zip-code level data are public-use, the actual CE zip-codes needed to link the data sets are highly confidential, so the analysis here was conducted by the authors at the Census Bureau and the Bureau of Labor Statistics. The CE does not generally receive address or zip code data from Census, in order to protect respondent confidentiality under Title 13. The tabulations for this paper were made on-site at BLS by the BLS staff sworn to uphold the provisions of Title 13.

<sup>16</sup> The data set covers units who span the 2007 and 2008 reporting years, so households sampled in 2009 first quarter (whose expenditures are measured for 2008 fourth quarter and for January and February 2009) are also included. Also consumer units with expenditures made in October, November and December 2006 are included as well since these data were collected in the first quarter of 2007 and could refer to 2006. Note that all 2007 income values (both CE income and zip-code level AGI) are inflated to 2008 dollars in order to combine the two years and increase the usable sample size.

being averaged to create the estimated response rates or the ratio of average CE income to average AGI.

### *Using Zip-Code Level AGI to Sort Households*

Using zip-code level AGI to proxy “true” income of non-respondents does raise a few concerns. First, the AGI concept itself is an imperfect measure of income, because it excludes non-taxable transfers along with other tax-free income such as municipal bond interest. The idea of non-taxable transfers usually evokes images of food stamps and other income maintenance programs, but it is probably more salient to note that for most Social Security recipients most or all of their Social Security is excluded from AGI. Thus, a retiree with \$20,000 in taxable pensions and \$20,000 in Social Security will show up with an AGI of \$20,000, even though the CE would identify them as having \$40,000 of income.

The second problem with using zip code-level mean AGI is the presence of dependent filers. As noted in the discussion of Table 2 in the previous section, the count of SOI “units” is much higher than CE consumer units or CPS households, because dependent children with income are required to file separate returns. Thus, the first two problems with using AGI—that AGI excludes non-taxable income and the averages include dependent filers—imply that average AGI for the zip code is a downward biased estimate of average household income. In the first case we are explicitly omitting important income components, and in the second we are splitting the household-level income across too many units. Indeed, the overall mean of CE income is about 20 percent higher than the mean of zip-code level AGI for the same households.

The third problem with using zip-code level AGI is that it excludes non-filers, but in this case there is no obvious bias in average AGI. Households who receive only non-taxable transfers

will not even show up in the SOI zip-code level data file, because they are not required to file tax returns.<sup>17</sup> Their exclusion from the zip-code file would reduce the total number of units, but would likely not change the income ranking of the zip-codes. That is, if a \$20,000 per year Social Security recipient lives in the same zip code as a \$20,000 per year wage earner, we would only observe the wage earner in the zip code AGI file, but the \$20,000 AGI would still be a good estimate of income for both households in the zip code.<sup>18</sup> Even if this is not an accurate assumption (see the next paragraph), the exclusion of non-filers is unlikely to affect the highest income zip-code areas, which are of most interest here.

The final problem with using zip-code level AGI is that zip-code may not be a narrow enough geographic classifier from a socioeconomic perspective, meaning there is significant income variation within zip codes. This potential problem motivates the second step of the approach implemented here, because in addition to looking for differences in response rates by AGI percentile, we also consider the ratio of CE respondent-reported incomes to average AGI. This second step is designed to capture differences in response by income within zip codes, and thus control for variations in within zip code incomes, especially at the top of the distribution where our attention is focused.

### *Response Rates by AGI Percentiles*

The first question addressed using the new zip-code linked data set is whether the probability of responding to the survey, when sampled, varies systematically with income.<sup>19</sup> All

---

<sup>17</sup> Note, however, the caveat mentioned above in footnote 12.

<sup>18</sup> One important direction for future research involves manipulating the CE to create an AGI concept and split the household into appropriate tax filing units. Although we cannot manipulate the SOI concept to match the CE, we can go in the other direction.

<sup>19</sup> Note that we are not testing whether or not the probability of being sampled varies with zip-code level income, though in principle that could be accomplished by comparing the sampled CE population against the entire SOI zip-code data set.

sampled CE units are assigned the average AGI for their zip-code, and the entire data set is sorted into 100 percentile groups (0th-1st, 1st-2nd, ..., 99th-100th). Although in principle this is a simple calculation, because response is a binary outcome, the analysis is complicated to some extent because it requires acknowledging the potential effects of existing BLS post-stratification (weighting) adjustments.

The simplest calculation involves the inverse of the raw sampling probability, which BLS refers to as BASEWT. The values for BASEWT in the CE are typically around 10,000, which means that a consumer unit in the sample represents 10,000 consumer units in the U.S. civilian non-institutional population—itsself plus 9,999 other consumer units that were not selected for the sample.<sup>20</sup> Using BASEWT, the simplest calculation of response by AGI involves taking the ratio of respondents (weighted by BASEWT) to sampled units (also weighted by BASEWT) within each AGI percentile (Figure 1, lowest set of markers).

The overall response rate across AGI percentiles is 74 percent for 2007 and 2008.<sup>21</sup> Figure 1 shows that the response rate for most AGI deciles is between 70 and 80 percent for most of the AGI distribution. Although the numbers exhibit a fair amount of variability, there is no clear pattern between (roughly) the 10<sup>th</sup> and 90<sup>th</sup> percentiles. The data do show lower response for the highest AGI percentiles, which confirms the hypothesized higher unit non-response for very-high income families. Overall, the response rate for the top five percentiles is 66 percent, and the top one percent by AGI has a response rate of 64 percent.

---

<sup>20</sup> There are some relatively minor adjustments to BASEWT that adjust for several types of operational and field sub-sampling. Examples of when sub-sampling is used include when a data collector visits a particular address and discovers multiple housing units where only one housing unit was expected or when more units are found in the listing than expected in rural areas that use an area frame.

<sup>21</sup> The fact that the BASEWT response rate of 74 percent exactly matches the response rates based on simple sample counts as noted earlier underscores the fact that the adjustments to BASEWT are empirically very small.

Interestingly, the response rates by AGI are higher than average at the bottom of the AGI distribution. The overall response rate based on BASEWT is 80 percent for the bottom five percentiles and 84 percent in the first percentile. Given the very large sample sizes involved in these calculations—over 800 sampled units in each AGI percentile—these higher response rates for lower income zip codes are noteworthy. Although we do not pursue an explanation for higher unit non-response by lower income households here, it is certainly an interesting area for further research.

Although the unadjusted response rates (based on BASEWT) suggest that higher income households are indeed under-represented in the CE respondent sample, there are two subsequent stages of BLS post-stratification that could remedy this under-representation.<sup>22</sup> The first-step involves the “non-interview” adjustment factor which involves applying differential adjustments based on estimated non-response patterns (this adjustment creates what BLS calls STAGE1WT). Specifically, this factor adjusts for interviews that cannot be conducted in occupied housing units due to a consumer unit's refusal to participate in the survey or the inability to contact anyone at the sample unit in spite of repeated attempts. This adjustment is performed separately for each month and “rotation group” (interview number) and yields 64 cells or factors based on region of the country, household tenure (owner or renter), consumer unit size, and race of the reference person.

If income is correlated with these 64 factors that affect unit non-response, then applying the non-interview adjustment factor could remedy the differential in response rates at very high (and very low) incomes. In fact, the correlation between zip-code level AGI and the BLS non-interview adjustment factor is positive and highly significant in our data. Nevertheless, as shown

---

<sup>22</sup> The discussion of CE weighting here largely follows the Bureau of Labor Statistics, *Handbook of Methods*, available on-line at [www.bls.gov/opub/hom/](http://www.bls.gov/opub/hom/).

in Figure 1 (middle set of markers) the adjustment factor raises response rates approximately uniformly across AGI percentiles. The overall adjustment factors are calibrated such that the adjusted overall response rate is basically 100 percent, meaning the new weights will sum to the count of originally sampled units, but nearly the same curvature in response rates at very high and very low percentiles is observed. That is, the *relative* response rates based on BASEWT is about 89 percent of the overall response rate, and this rises only to about 91 percent using STAGE1WT. Households in the top five percentiles are about 10 percent less likely to participate than the entire sample.

Finally, BLS applies a “calibration factor” that adjusts the weights to 24 “known” population counts to account for frame under-coverage. These “known” population counts are for age, race, household tenure (owner or renter), region, and urban or rural. The population counts are updated quarterly. Each consumer unit is given a calibration factor based on which of the 24 distinct groups they are in (this last adjustment creates FINALWT21, the weight that CE micro data uses are most familiar with).<sup>23</sup> Again, the calibration-adjusted response rates are shifted up versions of the STAGE1WT and BASEWT values (Figure 1, top set of markers) but there is no qualitative change in the pattern. As with BASEWT and STAGE1WT, the *relative* response rates in the top five percentiles are about 10 percent below the sample as a whole.

### *CE Incomes Relative to Average AGI*

The previous analysis demonstrated that there is a differential non-response in the very-high income AGI zip code areas. Although the CE income appears to be associated with zip-code level AGI, it is difficult to map these outcomes back to the univariate income distributions

---

<sup>23</sup> Note that there are infinitely many sets of calibration factors that make the weights add up to the 24 “known” population counts, and the CE selects the set that minimizes the amount of change made to the “initial weights” (initial weight = base weight x weighting control factor x noninterview adjustment factor).

shown earlier (Table 2) because CE households are being sorted by zip-code level AGI, not their own household income (which we cannot observe for non-respondents). The next part of the analysis provides more support for the proposition that the very highest income households are under-represented in the CE. For this purpose we use reported CE incomes, including the incomes imputed by BLS for consumer units who participate in the survey but who fail to respond to income questions.<sup>24</sup> In this second step, we compare average CE income to average AGI within each AGI percentile, and show that the ratio generally falls with income, and is dramatically lower at the top of the AGI distribution.

Across all AGI percentiles in the linked data set, mean CE income for respondents (based on FINALWT21) is about 20 percent higher than mean AGI for all sampled units (based on BASEWT, though the exact weight chosen does not affect this answer). However, there is a distinct downward pattern across AGI percentiles (Figure 2). The ratio of mean CE income to AGI is about 140 percent at the bottom of the income distribution, and falls steadily as AGI increases, before plummeting to 75 percent for the top two percentiles of AGI. Thus, Figure 2 complements Figure 1 in the following sense. Figure 1 shows that households in the top AGI percentile zip-codes are 10 percent less likely to participate than the rest of the sample, and Figure 2 suggests that the households within the top AGI percentiles that do participate are more likely to have lower incomes than the households in that zip-code who did not participate.

Although the conceptual differences between AGI and CE income make direct inferences impossible, it is worth noting that the combined insights from Figure 1 and Figure 2 probably go a long way towards explaining the income distribution differences presented in Table 2. For example, the CE finds about 7 percent fewer households above \$100,000 than the Survey of

---

<sup>24</sup> Exclusion of those “item non-respondents” would lower the CE’s overall income averages and totals. Imputation would have little effect on the Section 2 comparisons, because imputations are used in other surveys as well.



Consume Finances (SCF), which is similar in magnitude to the roughly 10 percent response differentials for the top five percentiles shown in Figure 1. Also, the ratio of average income in the CE to average income in the SCF for households above \$100,000 is 68 percent, which is in the same ballpark as the CE income to AGI ratios at the highest AGI percentiles. Although a direct mapping from the zip-code level AGI percentile analysis to univariate income distributions requires more research, the results here suggest that differential unit non-response probably goes a long way toward explaining the shortfalls.<sup>25</sup>

### *Probit Analysis*

An alternative approach to exploring the relationship between income and unit non-response involves estimating a binomial probit model, in which zip-code level AGI is included as a determinant of response status along with the 64-way matrix of stratifying variables used by the BLS in the weighting adjustment for non-response that creates STAGE1WT. Specifically, NR (in equation (1) below) is a binary variable that is equal to zero for responding CUs and one for those that did not participate in the survey. The regression also includes 63 dummy variables corresponding to all but one of the region-family size-race-housing tenure strata used for the non-response weighting adjustment in the CE. A third-order polynomial function in AGI is included using three variables: AGI,  $AGI^2/100$ , and  $AGI^3/10000$ .<sup>26</sup> The equation below is estimated using the same sample of 61,546 responding and 21,820 non-responding CUs described above, and observations are weighted by BASEWT.

---

<sup>25</sup> Perhaps the most crucial area for development involves reconciling the CE income and AGI concepts. Incorporating differential unit non-response into formal post-stratification adjustments requires eliminating any of the patterns in Figure 2 caused by a mismatch between the AGI and CE income concepts across AGI percentiles. That is, we know AGI is 20 percent higher on average, but there are good reasons to suspect the differential may vary with AGI percentile. See the discussion in the text about why AGI and CE income concepts diverge.

<sup>26</sup> As in the graphical analysis, AGI and CE income data are made more consistent by subtracting capital gains income from the former and food stamp benefits from the latter.

Each of the three AGI variables was asymptotically significant at the 0.01 percent significance level, even with all the stratifying variables held constant. A likelihood ratio test of the significance of the three AGI variables yielded a chi-square value of 170 with three degrees of freedom, which easily surpasses any usual significance level. The probit results of interest are:

$$(1) \text{ Probit (NR)} = [\text{dummy coefficients}] + \dots + 0.0050 \text{ AGI} - 0.0021 \text{ AGI}^2/100 + 0.0003 \text{ AGI}^3/10000$$

This equation implies a positive impact of zip code-level AGI on the nonresponse probability over the observed range, with the second derivative negative until the highest AGI levels, when it becomes positive.

The probit approach is indicative of how one might begin to think about creating an alternative to the BLS stage-one adjustments (STAGE1WT) using AGI along with the existing BLS stratifying variables. With the probit-based noninterview adjustments, the average adjusted response rate in the top five AGI percentiles is only about three percent below that of the sample as a whole, compared to about 10 percent using the BLS adjustments. By giving higher weights to CUs in higher-AGI areas, the probit approach does indeed imply higher aggregate weighted average CE incomes and expenditures, but the effects are modest.<sup>27</sup>

Using a revised weight based on the probit adjustment using AGI as an explanatory variable yields average income that is only about 0.35 percent higher and average spending that is about 0.22 percent higher than those using the BLS STAGE1WT. Relating this back to the

---

<sup>27</sup> It is important to recognize that if the BLS actually used these probit non-response adjustments it would necessarily lead to different calibration adjustments. The alternative calibration factors might be expected to reduce the differences between the current and probit-based income and expenditure estimates. Unfortunately, estimating new calibration factors was not feasible for this paper.

last section, the probit is able to capture the pattern shown in Figure 1, but not the pattern shown in Figure 2. That is, the simple adjustment can increase the weights of respondent households based on their AGI percentile, but it cannot capture the fact that the lower income households within zip-codes are the ones more likely to participate when sampled.<sup>28</sup>

#### **4. Why is Aggregate Consumer Expenditure Survey Spending So Low?**

If Personal Consumption Expenditures (PCE) in the National Income and Product Accounts (NIPA) are viewed as the truth about what consumers actually spend in a given time period, there are two possible high-level explanations for why aggregated spending in the Consumer Expenditure Survey (CE) is below the corresponding PCE totals. The evidence above provides some support for the first reason, which is that the very highest income households are under-represented. However, even if that “missing” income is added at the top of the income distribution, the aggregate spending mystery remains, and that would in any event actually deepen another mystery associated with the CE. The other mystery involves the second reason why CE spending is low: under-reporting of spending by at least some CE respondents.

The observed under-representation of very high-income households cannot fully explain the aggregate CE spending shortfall. The most extreme estimate of the aggregate CE income shortfall above \$100,000 comes from comparing the CE to the Survey of Consumer Finances (SCF). The SCF finds about \$1.7 trillion more income above \$100,000 than the CE, but if one

---

<sup>28</sup> Under contract with BLS, the Census Bureau is currently researching alternative variables to use in CE’s non-response and calibration adjustment processes. Income is one of the variables being considered. This research is addressing a number of questions, such as what variables are available for every household in the CE survey, both respondents and non-respondents; what qualities characterize “good” variables for these procedures; and what variables other surveys use. This research is expected to be completed in 2012. Given that the CE is designed to capture expenditures of all households and the results of this study that the very high income households are often not included in the sample, an oversampling strategy such as that employed by the SCF may also be worth considering. Implementation of oversampling could be expensive, and it would not by itself address a bias problem, but if combined with revised methods for non-response adjustment it could be a valuable improvement.

applies the BLS-reported ratio of expenditures to gross income for that group (61 percent) that implies total spending would rise by 16 percent, which explains perhaps half of the overall shortfall relative to PCE totals (as shown in Table 1).

Overall, published CE expenditures are lower than published CE after-tax incomes. For example, the ratio of published total expenditures to published after-tax income for CE respondents was 83 percent in 2006.<sup>29</sup> Given the relationship between aggregate spending and disposable income in the National Accounts data, that ratio probably should have been much higher.<sup>30</sup> Based on that aggregate perspective and the conclusion that misrepresented high income households only explains at most half of aggregate under-reporting, at least some of the shortfall in aggregate CE spending seems attributable to under-reporting of spending (given income) by at least some CE respondents.

Knowing that the overall spending to income ratio seems too low for the CE survey (based on comparisons to PCE) is a starting point, but it does not help with the distributional question of whether the propensity to under-report spending varies with income itself. Researchers interested in using the CE for distributional analysis of questions about topics like consumption-expenditure versus income inequality, saving rates, or the distributional burden of consumption taxes, rely completely on the empirical joint distribution of expenditures and income. If the problem is proportional under-reporting of expenditures for all CE respondents,

---

<sup>29</sup> These calculations are based on published BLS numbers, even though the reported values have both conceptual problems and systematic reporting errors in at least one key variable. Conceptually, for example, BLS counts Social Security taxes and employee contributions to pensions as expenditures, but they do not count mortgage principal repayments as spending. For these and other reasons the concept of after-tax income minus expenditures is not in any sense a pure “saving” estimate, but there are biases in both directions, and fixing those would require unavailable information such as net home equity extraction needed to measure net mortgage principal payments. There are also some measurement biases in the table that BLS is aware of and working on—for example, based on comparison of effective tax rates with other sources, under reporting of income taxes could account for several percentage points of the overall cash-flow discrepancy, and even more for higher-income respondents.

<sup>30</sup> See, for example, Bosworth, Burtless, and Sabelhaus (1991) for a discussion of what is involved with reconciling aggregate and household-level saving concepts.

then the simple solution is to scale up spending for all households (perhaps by type of spending) before undertaking any distributional analysis (see Slesnick (2001) and Meyer and Sullivan (2011) for a similar approach). However, if the propensity to under-report rises with spending (and thus with income) then some sort of differential adjustments are warranted.

The estimated pattern of spending to income ratios by income in the CE may have flaws, but if it does, those flaws are not a new phenomenon (Figure 3). A comparison of published BLS data for 1972-73, 2003, and 2010 shows that the ratio of spending to unadjusted after-tax income at any given level of income has not changed much in 40 years.<sup>31</sup> Overall, the ratio of spending to after-tax income fell from 89 percent in 1972-73 to 84 percent in 2003 and 79 percent in 2010.<sup>32</sup> Based on aggregate trends, the overall spending to income ratio should have been higher in the last two periods than in the first. However, it is difficult to see differential declines in spending to income ratios across income groups. The first-order differences in spending to income ratios occur across income groups at each point in time, not across time periods.<sup>33</sup>

The ratios of total expenditure to after-tax incomes by income shown in Figure 3 exhibit a dramatic pattern, and although there are some conceptual issues and systematic reporting errors with income taxes in the BLS tabulations, those sorts of corrections do not fundamentally change that pattern.<sup>34</sup> The ratio of spending to income at low income levels seems implausibly high, and

---

<sup>31</sup> Each point on the chart marks average total expenditures divided by average after-tax income, at the value of after-tax income reported in the BLS tables. Values average after-tax income in 1972-73 and 2003 are inflated to 2010 dollars. The year 2003 marks the first year in which BLS published “high income” tables for the modern (post-1980) on-going CE survey.

<sup>32</sup> The overall ratio of spending to after-tax income fell between 1972-72 and the latter two periods, even though the ratios across income groups were stable, because more households moved into higher real income groups with much lower spending to after-tax income ratios. Thus, given what we know about aggregate saving, the ratios should have shifted (or twisted) up.

<sup>33</sup> The stability of spending to income ratios across income groups also raises concerns about the approach used by Aguiar and Bils (2011) to “correct” for bias in studies that compare consumption versus income inequality. They use the 1972-73 CE survey to estimate Engel curves, and impute missing spending in the 1980s based on those estimated relationships and an aggregate scaling factor. If under-reporting for higher income families was just as bad in the 1970s as it is today, then they are effectively just inflating observed spending to match aggregates.

<sup>34</sup> See footnote 30 above.

the ratio of spending to income at the top seems implausibly low. There are most likely problems with both income and expenditure reporting, and sorting households by income simply highlights those errors.

In any household survey there will be measurement error, and given that the CE is focused on spending rather than income, it is not surprising that income may be poorly reported for some households.<sup>35</sup> The households who under-report income are (by definition) more likely to be sorted into the bottom of the income distribution, and thus the high ratios of expenditure to income at low incomes is not surprising. The argument that income is missing at the bottom is reinforced by a pragmatic view of lower-income households. It is impossible to spend twice your income (Figure 3) if you have no assets to draw down and no access to credit, which is the basic conclusion one takes away from wealth surveys like the SCF or Panel Survey of Income Dynamics. Thus, except for students, households with temporary business losses, and retirees drawing down assets, the high rates of implied dissaving by lower income households in the CE are already implausible, and proportional scaling up of spending would only make those ratios more implausible.

Although proportional scaling up under-reported expenditures at the bottom of the income distribution seems to worsen one existing mystery about spending to income ratios, the possibility that spending is more likely to be under-reported by higher-income (and thus higher-spending) families is consistent with observed patterns. In the same sense that lower-income households cannot, on average, spend twice what they earn, it is unrealistic to think that families

---

<sup>35</sup> Indeed, the CE data includes a number of consumer units who either refuse to answer or say they don't know, which is why income is imputed for a significant number of cases. The CE imputation procedures, described in Passero (2009) and Paulin and Ferraro (1996), focus on preserving the consumption to income relationship for those households who do participate, by using expenditures as an explanatory variable in the imputation procedures. The conclusions of this paper might suggest some reconsideration of the current imputation procedures to reflect non-random nonresponse.

above \$100,000, on average, save the fraction of their disposable income implied by Figure 3, using it for purchasing stocks, bonds, and other investments that are not captured by the CE. If that were the case, average wealth to income ratios for higher income households would quickly explode, and they would be much different than what we observe in actual wealth surveys.<sup>36</sup>

## 5. Conclusions

Only the very highest income households seem to be under-represented in the Consumer Expenditure Survey (CE), and the mystery of overall under-reported spending in the CE is not fully explained by that shortcoming. At least some of the shortfall in aggregate CE spending seems attributable by under-reported spending by at least some CE respondents, and that has implications for research that relies on the relationship between spending and income in micro data. The observation that spending to income ratios fall with reported income in the CE implies that consumption-expenditure inequality will be less than income inequality, and the extent to which this ratio falls with income (and changes over time) has a dramatic impact on the estimated relationship between consumption-expenditure and income inequality. Also, if this pattern in the spending-to-income ratios is partially due to measurement of total spending, then the amount of dissaving at low incomes and saving at high incomes will both be exaggerated, and consumption taxes will appear (perhaps wrongly) to be highly regressive alternatives to income taxes.

---

<sup>36</sup> Some might argue that these simple calculations ignore income fluctuations, because households do not stay in the same income group from one year to the next. That is exactly the argument addressed by Sabelhaus and Groen (2000) who use data on income mobility from the PSID to test whether movements across income groups can explain the pattern of consumption to income in the CE. The answer they find is clearly no—there is not enough income mobility, even under the most extreme assumptions about consumption smoothing.

Resolving whether expenditures are proportionally under-reported for all CE respondents or disproportionately for higher income (and thus higher spending) respondents is a crucial task facing the current CE redesign effort. It may be the case that the demands placed on respondents in the current CE are simply too daunting, because respondents are asked to remember several hundred spending items for each month in a three-month recall period. Thus, one approach to reconciling the difference between incomes and spending across income groups might involve streamlining the collection of spending totals, so that even high spenders will have a better chance to accurately estimate and report their total spending.<sup>37</sup>

## 6. References

- Aguiar, Mark A., and Mark Bilz. 2011. "Has Consumption Inequality Mirrored Income Inequality?" National Bureau of Economic Research, Inc, NBER Working Papers: 16807. (February)
- Attanasio, Orazio, Gabriella Berloffa, Richard Blundell, and Ian Preston. 2002. "From Earnings Inequality to Consumption Inequality," *Economic Journal*, 112(478): 52-59. (March)
- Blair, Caitlin. 2011. "Constructing a PCE-Weighted Consumer Price Index," prepared for the December, 2011 CRIW/NBER Conference, *Improving the Measurement of Consumer Expenditures*.
- Bosworth, Barry, Gary Burtless, and John Sabelhaus. 1991. "The Decline in Saving: Evidence from Household Surveys," *Brookings Papers on Economic Activity*, 1991(1):183-241.
- Browning, Martin, and Thomas Crossley. 2009. "Are Two, Cheap Noisy Measures Better Than One Expensive, Accurate One?" *American Economic Review*, 99(2):99-103. (May)
- Bucks, Brian K., Arthur B. Kennickell, Traci L. Mach, and Kevin B. Moore. 2009. "Changes in U.S. Family Finances from 2004 to 2007: Evidence from the Survey of Consumer Finances," *Federal Reserve Bulletin*, 1st Quarter 2009, 95:A1-55.
- Burkhauser, Richard V, Shuaizhang Feng, Stephen P. Jenkins, and Jeff Larrimore. 2009. "Recent Trends in Top Income Shares in the USA: Reconciling Estimates from March CPS and IRS Tax Return Data." National Bureau of Economic Research, Inc, NBER Working Papers: 15320. (September)

---

<sup>37</sup> Browning and Crossley (2009) discuss the merits of collecting aggregated versus disaggregated spending data.



Crossley, Thomas F. 2009. "Measuring Consumption and Saving: Introduction," *Fiscal Studies*, 30(3-4): 303-07. (September-December)

Garner, Thesia I., and Kathleen Short. 2009. "Accounting for Owner-Occupied Dwelling Services: Aggregates and Distributions," *Journal of Housing Economics*, 18(3):233-48. (September)

Garner, Thesia I., George Janini, William Passero, Laura Paszkiewicz, and Mark Vendemia. 2006. "The CE and the PCE: A Comparison," *Monthly Labor Review*, 129(9):20-46. (September)

Groves, Robert M. 2006. "Nonresponse Rates and Nonresponse Bias in Households Surveys," *Public Opinion Quarterly*, 70:5, pp. 646-675 (December).

Heathcote, Jonathan, Fabrizio Perri, and Giovanni L. Violante. 2010. "Unequal We Stand: An Empirical Analysis of Economic Inequality in the United States, 1967-2006," *Review of Economic Dynamics*, 13(1):15-51. (January)

Johnson, David, and Stephanie Shipp. 1997. "Trends in Inequality Using Consumption-Expenditures: The U.S. from 1960 to 1993," *Review of Income and Wealth*, 43(2):133-52. (June)

Kennickell, Arthur B., and Louise R. Woodburn. 1999. "Consistent Weight Design for the 1989, 1992 and 1995 SCFs, and the Distribution of Wealth," *Review of Income and Wealth*, 45(2):193-215. (June)

Kennickell, Arthur B. 2009. "Ponds and Streams: Wealth and Income in the U.S., 1989 to 2007," Board of Governors of the Federal Reserve System, Finance and Economics Discussion Series: 2009-13.

King, Susan L., Boriana Chopva, Jennifer Edgar, Jeffrey M. Gonzales, Dave E. McGrath, and Lucilla Tan. 2009. "Assessing Nonresponse Bias in the Consumer Expenditure Interview Survey," *Section on Survey Research Methods*, Joint Program on Survey Methods Annual Conference, 1808-1816.

Krueger, Dirk, Fabrizio Perri, Luigi Pistaferri, and Giovanni L. Violante. 2010. "Cross-Sectional Facts for Macroeconomists," *Review of Economic Dynamics*, 13(1): 1-14. (January)

Krueger, Dirk, and Fabrizio Perri. 2006. "Does Income Inequality Lead to Consumption Inequality? Evidence and Theory," *Review of Economic Studies*, 73(1):163-93. (January)

Meyer, Bruce D., Wallace K.C. Mok, and James X. Sullivan, 2009. "The Under-Reporting of Transfers in Household Surveys: Its Nature and Consequences." National Bureau of Economic Research, Inc, NBER Working Papers: 15181, originally published September 2006, revised July 2009.

Meyer, Bruce D. and James X. Sullivan, 2011. "The Material Well-Being of the Poor and the Middle Class Since 1980," American Enterprise Institute Working Paper #2011-44.

McCully, Clinton P., Brian C. Moyer, and Kenneth J. Stewart. 2007. "Comparing the Consumer Price Index and the Personal Consumption Expenditures Price Index," *Survey of Current Business*. (November)

Passero, Bill. 2009. "The Impact of Income Imputation in the Consumer Expenditure Survey," *Monthly Labor Review*, 132(8):25-42. (August)

Paulin, Geoffrey D., and David L. Ferraro. 1996. "Do Expenditures Explain Income? A Study of Variables for Income Imputation," *Journal of Economic and Social Measurement*, 22(2):103-28.

Sabelhaus, John, and Jeffrey A. Groen. 2000. "Can Permanent-Income Theory Explain Cross-sectional Consumption Patterns?" *Review of Economics and Statistics*, 82(3):431-38. (August)

Short, Kathleen\_et al. 1998. "Poverty-Measurement Research Using the Consumer Expenditure Survey and the Survey of Income and Program Participation," *American Economic Review*, 88(2):352-56. (May)

Slesnick, Daniel, 2001 *Consumption and Social Welfare*, Cambridge: Cambridge University Press.

U.S. Department of Labor, Bureau of Labor Statistics. *Handbook of Methods*. Available on-line at [www.bls.gov/opub/hom/](http://www.bls.gov/opub/hom/).

**Table 1. Ratio of Consumer Expenditure Survey Aggregates to Comparable NIPA Personal Consumption Expenditure Measures**

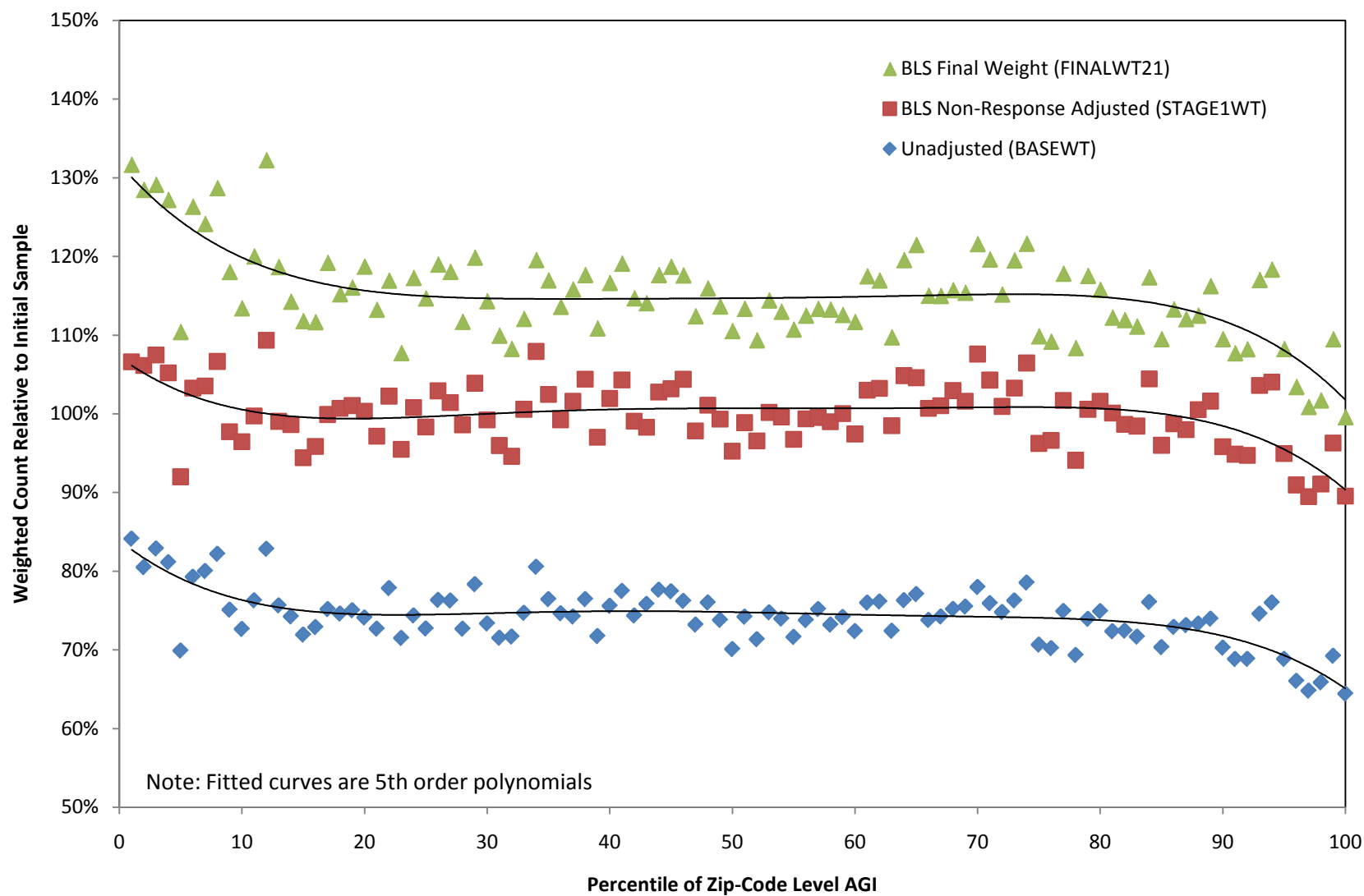
<b>Ratio of CE to PCE for Comparable Categories</b>					
<b>Year</b>	<b>All Goods and Services</b>	<b>Durable Goods</b>	<b>Non-Durable Goods</b>	<b>Owned Housing</b>	<b>Other Services</b>
<b>Garner, McClelland, and Passero (2009)</b>					
1992	0.88	0.88	0.69	1.23	0.90
1997	0.88	0.80	0.67	1.26	0.86
2002	0.84	0.75	0.63	1.25	0.82
2003	0.82	0.79	0.61	1.26	0.80
2005	0.83	0.75	0.63	1.26	0.81
2007	0.81	0.69	0.61	1.30	0.81
<b>BLS Published Estimates Based on Latest NIPA Crosswalk</b>					
2003	0.77	0.68	0.62	1.18	0.70
2005	0.79	0.68	0.64	1.16	0.73
2007	0.78	0.61	0.63	1.22	0.71
2009	0.78	0.60	0.64	1.11	0.72

**Table 2. Income Distribution in the Consumer Expenditure Survey and Three Other Data Sets, 2006**

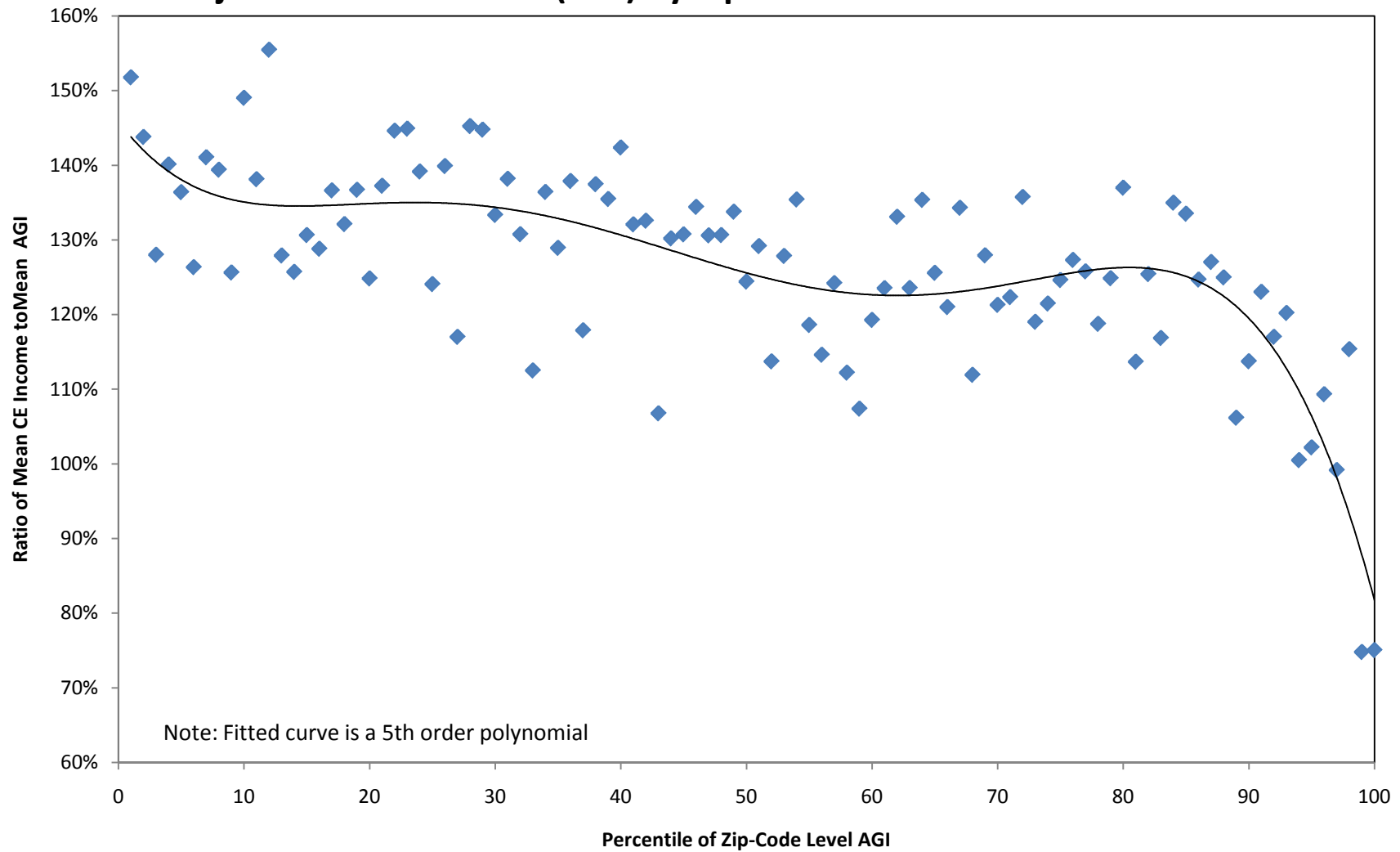
	Income Category			
	Less than \$50,000	\$50,000 to \$99,999	\$100,000 or More	All Incomes
<b>Consumer Expenditure Survey</b>				
Number of Units (Millions)	65.1	34.9	18.9	118.8
Total Income (Billions)	\$1,608	\$2,476	\$3,111	\$7,195
<b>Differences from Consumer Expenditure Survey</b>				
<b>Current Population Survey</b>				
Number of Units (Thousands)	-5.5	-0.6	3.2	-2.8
Total Income (Billions)	-\$104	-\$55	\$688	\$528
Total Income (Percent)	-6.5%	-2.2%	22.1%	7.3%
<b>Survey of Consumer Finances</b>				
Number of Units (Thousands)	-1.6	-2.7	1.5	-2.7
Total Income (Billions)	-\$8	-\$170	\$1,832	\$1,654
Total Income (Percent)	-0.5%	-6.9%	58.9%	23.0%
<b>Statistics of Income</b>				
Number of Units (Thousands)	27.2	-4.9	-2.8	19.6
Total Income (Billions)	\$191	-\$357	\$1,002	\$836
Total Income (Percent)	11.9%	-14.4%	32.2%	11.6%

Notes: SCF and SOI income exclude capital gains.

**Figure 1. Consumer Expenditure Survey (CE) Response Rates  
by Zip-Code Level Adjusted Gross Income (AGI) Percentile**



**Figure 2. Ratio of Mean Consumer Expenditure Survey (CE) Income to Adjusted Gross Income (AGI) by Zip-Code Level AGI Percentile**



**Figure 3. Expenditure to After-Tax Income Ratios in Published CE data**

