

## Income imputation and the analysis of consumer expenditure data

*The Consumer Expenditure Survey now provides imputed income data from 2004 forward for households that do not report a specific income value; an examination of how income imputation affects the analysis of expenditure data shows that the results most sensitive to imputation are statistics that focus on households with lower levels of expenditures*

Jonathan D. Fisher

**T**he Bureau of Labor Statistics (BLS) Consumer Expenditure Survey (CE) began imputing income in its 2004 data. Imputation predicts income for households that reported receiving income but failed to report a specific value. Many national household surveys such as the Current Population Survey and the Survey of Consumer Finances impute missing income values. While imputation is common practice, researchers should take some precautions when using imputed data.

This article examines how income imputation affects analysis of the CE *expenditure* data. Most importantly, researchers who use both income and expenditures data from 2004 forward no longer have to restrict their sample to households that reported income. This study presents results for the restricted sample employed before imputation was introduced and compares them with results using the sample that should be employed after imputation. The study also compares the distribution of expenditures and measures of well-being—such as the Gini coefficient and the poverty rate—in the two samples.

The other large effect of adopting income imputation is that there may be a break in time series data that use multiple years of CE data. Because BLS will only provide imputed income data from 2004 forward, researchers who want to create a time series using income and expenditures will not have imputed income data for the period before 2004. This study uses data from 2002 to 2004 to show how the introduction of income imputation creates

a break in the time series for some statistics (such as the poverty rate).

The data section of the article describes the relevant factors of the CE, and the imputation section provides an overview of the imputation procedure and how it interacts with expenditures. The methodology section describes the sample, defines the measure of expenditures, and defines other key variables. It then compares the distributions of expenditures before and after imputation and looks at how measures of well-being are affected by the introduction of income imputation.

### Data

The CE Interview Survey is a continuing quarterly survey of consumer units. A consumer unit consists of members of a household who are related or share at least two of the three major expenditures: housing, food, and other living expenses. In each consumer unit, one individual is referred to as the reference person, who is the person who rents or owns the residence as designated by the respondent. Data are collected from consumer units and the individuals within these consumer units five times over a 13-month period. The first interview is used for bounding purposes and is not released publicly. The remaining four quarters of data are released publicly, and these are the data used in this analysis. By restricting the sample to consumer units that appear in all four quarterly interviews, a measure of yearly expenditures for each consumer unit can be created.

Jonathan D. Fisher is an economist formerly in the Division of Price and Index Number Research, Bureau of Labor Statistics. E-mail: econofish@gmail.com

The CE has 18 income variables. The following 6 variables are collected for each *individual* in the consumer unit: wages and salaries, self-employment income (nonfarm), farm income, Social Security benefits, railroad retirement benefits, and Supplemental Security Income benefits. The remaining 12 variables are collected for the *consumer unit as a whole*: pension income, interest income, dividend income, royalty income, unemployment benefits, workers' compensation benefits, child support, alimony, income from roomers or boarders, income from other rental units, food stamp benefits, and other income.

BLS creates a complete income reporter designation to determine whether consumer units provided sufficient income data for use in official publications.<sup>1</sup> A consumer unit is designated a complete income reporter if it meets one of the following three criteria:

1. The reference person reports a nonzero amount for a major income source. BLS defines major sources of income as wage and salary, self-employment, farm income, Social Security benefits, railroad retirement benefits, or Supplemental Security Income benefits.
2. At least one other consumer unit member reports a nonzero amount for a major income source and reports valid zeros for all major income sources for the reference person.
3. The consumer unit reports a nonzero amount for at least one other income source and valid zeros for all major sources for all members.

A consumer unit could be classified as a complete income reporter and still not provide a full accounting of its income. For example, the reference person could report wage and salary income but fail to report a valid amount for its alimony income; this consumer unit would be classified as a complete income reporter under condition (1). In 2004, 87 percent of consumer units were complete income reporters and only 64 percent of complete income reporters reported valid amounts for every income source. Overall, 44 percent of consumer units in 2004 reported an invalid amount for at least 1 of the 18 income components.<sup>2</sup>

## Income imputation

For the 44 percent of consumer units that reported at least one invalid income amount, BLS imputes income using the multiple imputation technique described by D. B. Rubin in his 1987 book *Multiple Imputation for Nonresponses in Surveys*.<sup>3</sup> As implemented by BLS, multiple imputation is an iterative regression-based approach, where the data for the regression comes from the valid non-zero reporters for each income component. The dependent variable equals the income component being imputed, and each of the 18 components is imputed

separately. The independent variables include demographic characteristics of the consumer unit and a variable that equals the quarterly expenditure outlays for the consumer unit.<sup>4</sup>

An initial regression is run with all of the independent variables. Any independent variable that is not statistically significant at the 15-percent confidence level using a two-sided test is removed from the regression model. A second regression is then run with the variables that were statistically significant in the initial regression. This iterative process continues until all remaining variables are statistically significant at the 15-percent level. The coefficients from this last regression are used to predict income for invalid reporters.<sup>5</sup> Through this iterative removal of independent variables, the quarterly expenditure-outlays variable may or may not be in the final regression for a given income variable. If the expenditure outlays variable remains in the regression model, the level of quarterly expenditure outlays affects imputed income. This creates dependence between expenditure outlays and income that may affect conclusions about the relationship between expenditures and income.

Imputed income also directly affects the official BLS measure of expenditures. The publication category "personal insurance and pensions" includes employee contributions to Social Security that are derived from the wage and salary variable. In the 2004 sample used in this study, 25 percent of consumer units had wage and salary income imputed. After imputing missing wage and salary data for these individuals, BLS then assumes that each individual paid Social Security taxes at the required 7.65 percent rate.<sup>6</sup> This 7.65 percent is added to the official personal insurance and pensions category and in the official BLS measure of total expenditures. Because Social Security taxes are also imputed, in its publications that use imputed income, BLS warns that "average annual expenditures and expenditures for personal insurance and pensions are not strictly comparable" to data from previous years.<sup>7</sup>

## Methodology

The expenditure variable used in this study differs from the official BLS measure of total expenditures. The personal insurance and pensions category is excluded from total expenditures under the assumption that such expenditures are more accurately classified as savings or taxes. The measure of consumption expenditures used here equals the sum of outlays for housing, food, transportation, apparel, medical care, entertainment, gifts, and miscellaneous items. (See the appendix for additional details about consumption expenditures.) The definition follows much of the literature that defines expenditures as outlays.<sup>8</sup> However, there is no consensus about the proper definition of outlays or expenditures. J. M. Rogers and M. Gray, for example, define three measures of expenditures, and all of the studies referenced in this article deviate from those three definitions. The definition employed here also differs slightly from that of Rogers and Gray

because they include insurance and pension contributions in their expenditure outlays measure while this study excludes them.

Two samples of consumer units are employed to show how income imputation affects the analysis of consumption expenditures. The first sample consists of complete income reporters as defined previously. This is the sample often used by past researchers who studied income and expenditures together. But income imputation allows researchers to utilize *all* consumer units. Thus, the second sample used in this study includes both complete and incomplete income reporters.<sup>9</sup>

One might be concerned that adding incomplete reporters may alter the sample in observable ways. Research that excludes incomplete income reporters implicitly assumes that incomplete reporters represent a random sample of the population, which suggests that the inclusion of them would not affect the sample. The purpose of this study is to show how the inclusion of incomplete reporters might affect the conclusions drawn about the distribution and the analysis of consumption expenditures after income imputation. It also looks at how a time series of economic statistics might be affected by the introduction of income imputation.

The study begins by comparing the distribution of consumption expenditures for the two samples by presenting the percentiles, the Gini coefficient, and the poverty rate for each. For the poverty rate, first, the level of consumption expenditures is compared with the official poverty thresholds. Then, con-

sumption expenditures are used in an Engel curve regression to further test the sensitivity of results from the two samples. An Engel curve estimates how the share of food expenditures in total consumption expenditures depends on permanent income and other factors. Previous studies have estimated Engel curves for all households, Hispanic households, and for poor households.<sup>10</sup> The dependent variable equals the share of total consumption expenditures devoted to food. The key independent variable is the log of permanent income. Much of the earlier research uses consumption expenditures as a proxy for permanent income. This study follows that methodology and includes the log of consumption expenditures and its square. The regression also contains a number of control variables—the number of adults in the household, the number of children in the household, and the square for each—as well as dummy variables for region, education, race, and year.

## Comparing distributions

Most of the analysis in this article compares two samples: all consumer units and complete income reporters. Focusing on these two groups obscures some of the differences in the data because incomplete income reporters are a small proportion of all consumer units. Incomplete reporters are only 13.6 percent of all consumer units. Tables 1 and 2 and chart 1 include separate results for incomplete income reporters.

**Table 1. Demographic characteristics by income reporter status, 2002–04**

	Complete income reporters	Incomplete income reporters	All consumer units
Number of observations .....	11,271	1,780	13,051
Age of reference person (in percent)			
Age 25 or less .....	4.1	3.5	4.0
Age 26 to 35 .....	15.1	12.9	14.8
Age 36 to 45 .....	21.6	23.1	21.9
Age 46 to 55 .....	21.4	23.5	21.7
Age 56 to 65 .....	15.3	17.3	15.5
Age 66 and older .....	22.5	19.7	22.1
Family size (mean)			
Adults .....	1.8	1.8	1.8
Children .....	.8	.8	.8
Family type (in percent)			
Single, no children .....	26.6	26.1	26.5
Single parent .....	5.5	4.6	5.4
Married, no children .....	23.7	24.3	23.8
Married with children .....	26.9	27.0	26.9
Married, other .....	4.3	4.7	4.3
Other family type .....	13.0	13.3	13.1
Education of reference person (in percent)			
High school dropout .....	15.6	13.8	15.4
High school graduate .....	27.8	29.6	28.0
Some college .....	28.8	30.2	29.0
College graduate .....	27.8	26.4	27.6
Race of reference person (in percent)			
White .....	84.5	81.0	84.0
Black .....	10.5	15.2	11.1
Other race .....	5.0	3.8	4.9
Region of residence (in percent)			
Northeast .....	19.1	20.1	19.2
West .....	20.9	18.9	20.6
South .....	36.0	34.1	35.8
Midwest .....	24.0	26.9	24.4
Percent who live in urban areas .....	72.4	71.3	72.3

SOURCE: Consumer Expenditure Interview Survey (2002–2004). All data are population weighted.

**Table 2. Distribution of consumption expenditures by income reporter status, 2002–04**

Item	Complete income reporters	Incomplete income reporters	All consumer units
Mean .....	35,441	31,099	34,845
10th percentile .....	11,899	9,973	11,568
25th percentile .....	18,870	15,614	18,386
50th percentile .....	29,542	25,905	28,953
75th percentile .....	45,060	40,093	44,436
90th percentile .....	64,577	58,543	63,703
Inequality measure			
Gini .....	.32	.32	.32

SOURCE: Consumer Expenditure Interview Survey (2002–2004). All data are population weighted. Expenditure data are in real 2004 dollars, adjusted using the CPI-U-RS.

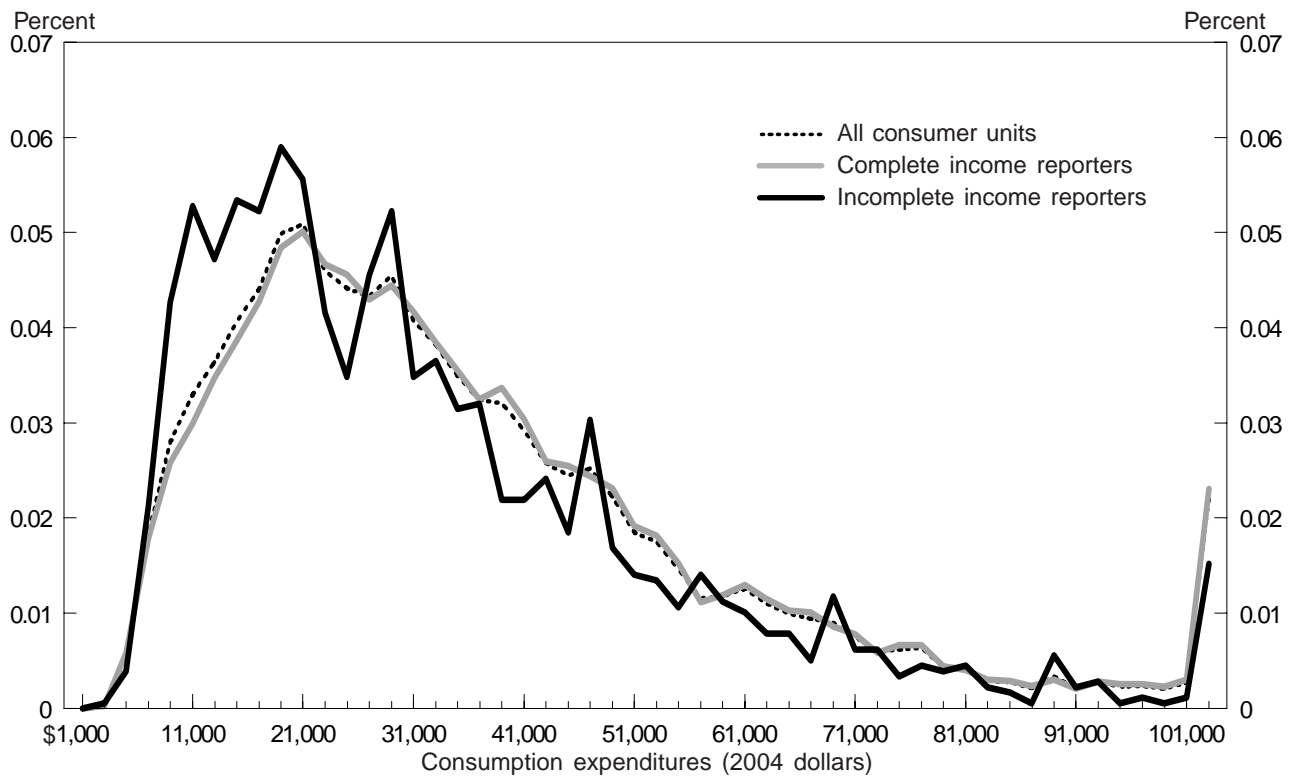
Table 1 compares selected demographic characteristics for complete income reporters, incomplete income reporters, and all consumer units. The mean age of the reference person for each group rounds to 51.5 years. As the age brackets shown in table 1

suggest, however, the means mask some heterogeneity in the distribution of age. Incomplete income reporters are less likely than complete reporters to be under age 35 and over age 65.<sup>11</sup> Family size is identical across samples, with each sample having 1.8 adults and 0.8 children, on average. There are also only small differences across samples for family type, education, region of residence, and urban status. There are noticeable differences in race, with incomplete reporters more likely to be black and less likely to be white.

Table 2 moves to the consumption expenditures data and presents percentiles of the consumption expenditures distribution. The means and medians are presented as measures of central tendency. The mean exceeds the median by approximately \$6,000 for complete and incomplete reporters, indicating that the distributions are right skewed. Chart 1 shows the extent of the skew in more detail by presenting the frequency distribution of consumption expenditures for complete income reporters, all consumer units, and incomplete income reporters. All three distributions exhibit a long right tail.

Table 2 and chart 1 also reveal that consumption expenditures are higher for complete reporters across all percentiles of the

**Chart 1. Frequency distribution for consumption expenditures, 2002–04**



SOURCE: Consumer Expenditure Interview Survey. All data are population weighted.

consumption expenditures distribution.<sup>12</sup> Incomplete reporters are more likely to be in the lower half of the consumption expenditures distribution. The comparison of interest is between complete reporters and all consumer units because these are the two samples researchers will tend to use. In addition, because incomplete reporters are a small fraction of the sample for all consumer units, the distributions for complete reporters and all consumer units virtually overlap at levels of consumption expenditures greater than \$23,000. The large grouping of incomplete reporters with consumption expenditures of less than \$23,000 causes the distributions of complete reporters and all consumer units to differ visibly in this range.

An additional way to describe a distribution is to look at the level of inequality, which can be seen in the Gini coefficients shown at the bottom of Table 2.<sup>13</sup> The Gini coefficient for complete income reporters equals the Gini for all consumer units. The two Gini coefficients differ only in the third digit after the decimal place.

Another measure economists often focus on is the poverty rate.<sup>14</sup> Table 3 presents the percent of consumer units with consumption expenditures below the official U.S. poverty threshold, as designated by the Census Bureau. The table shows that 10.8 percent of complete reporters had consumption expenditures below the poverty threshold, while 11.5 percent of all consumer units were consumption expenditure poor. With over 102 million consumer units represented by the 2004 CE, this difference in the poverty rate of 0.7 percentage point means that an additional 748,000 consumer units are considered consumption expenditure poor in the sample of all consumer units. Moreover, this difference persists when the sample is split by family type and race.

**Table 3. Percent of consumer units below official poverty threshold using consumption expenditures, 2002–04**

Item	Complete income reporters	All consumer units
All .....	10.8	11.5
By family type .....		
With children in consumer unit .....	10.7	11.3
Elderly head of household .....	16.3	17.4
By race of the reference person .....		
White .....	8.9	9.4
Black .....	23.3	26.9
Other race .....	12.5	12.7

SOURCE: Consumer Expenditure Interview Survey (2002–2004). All data are population weighted. Expenditure data are in real 2004 dollars, adjusted using the CPI-U-RS. The poverty thresholds are the official thresholds published by the U.S. Census Bureau.

## Time series analysis

Because of the difference in the poverty rate between the two samples, the introduction of income imputation could affect conclusions about the change in the poverty rate between 2003 and 2004. A researcher that creates a time series of poverty rates might use complete income reporters before 2004 and all consumer units in 2004. Table 4 shows that the poverty rate for complete reporters equals 10.6 percent in 2003 and 10.9 percent in 2004. If all consumer units are used for 2004, then the poverty rate equals 11.9 percent. Thus, depending on which sample is used in 2004, the poverty rate increased by either 0.3 percentage point or 1.3 percentage points from 2003 to 2004, a large difference that should be taken into account in analysis.

A similar issue might arise for the Gini coefficient and other statistics of interest. Using complete reporters, the results suggest that the Gini coefficient for consumption expenditures increased by 3.8 percent between 2003 and 2004. (See table 4.) But if complete reporters are used in 2003 and all consumers are used in 2004, inequality increased by 4.1 percent.

Table 4 also presents the change in consumption expenditures between 2003 and 2004 at different points of the consumption expenditures distribution. The growth in mean consumption expenditures between 2003 and 2004 equals 0.6 percent when using complete income reporters in both years, while it equals –1.7 percent using complete reporters in 2003 and all consumer units in 2004. The change in expenditures is always lower when using all consumer units in 2004 than when using complete income reporters in 2004. Taken as a whole, Table 4 suggests that researchers using multiple years of CE data must be aware of the fundamental change in the sample between 2003 and 2004. Statistics that focus on the bottom of the consumption expenditures distribution, such as the poverty rate, will be most sensitive to the change in the sample. Other statistics that consider the whole distribution, like the Gini coefficient, may not be as sensitive to the change.

## Regression analysis

As a final test of the sensitivity of the consumption expenditures data to the introduction of income imputation, the data are analyzed in a regression framework. As described in the methodology section, Engel curves are estimated, which means that the share of food is regressed on consumption expenditures and other control variables. Table 5 displays ordinary least squares estimates of the coefficient on consumption expenditures and its square. The coefficients match expectations for complete income reporters and all consumer units; the linear term on consumption expenditures is negative and statistically significant, while the quadratic term is positive, statistically significant, and small relative to the linear term.<sup>15</sup>

**Table 4. Examining the change in consumption expenditures from 2003 to 2004**

Item	Complete income reporters		All consumer units	Percent change	
	2003	2004	2004	Complete income reporters	Complete income reporters to all consumer units
	(A)	(B)	(C)	[(B) - (A)] / (A)	[(C) - (A)] / (A)
Poverty rate .....	10.64	10.94	11.85	2.8	11.3
Gini coefficient .....	.32	.33	.33	3.8	4.1
Mean .....	23,143	23,291	22,739	.6	-1.7
25 <sup>th</sup> percentile .....	13,470	13,124	12,666	-2.6	-6.0
50 <sup>th</sup> percentile .....	19,899	19,534	19,022	-1.8	-4.4
75 <sup>th</sup> percentile .....	28,799	28,565	28,055	-8	-2.6

SOURCE: Consumer Expenditure Interview Survey (2002–2004). All data are population weighted. Expenditure data are in real 2004 dollars using the CPI-U-RS. The poverty thresholds are the official thresholds published by the U.S. Census Bureau.

Next, the sample is restricted to those households that have consumption expenditures below the official poverty threshold. Restricting it to low consumption expenditure consumer units allows for testing the sensitivity of the regression results in the portion of the sample most likely to be affected by income imputation. Instead of a negative coefficient on consumption expenditures, the coefficient is positive and statistically significant.<sup>16</sup> While the coefficient on consumption expenditures is positive and significant for both complete reporters and all consumer units, the point estimate for the linear term for all consumer units is 41 percent higher. Thus, the regression results are sensitive to the sample chosen, especially when the sample comes from the bottom part of the consumption expenditures distribution.

THE CONSUMER EXPENDITURE SURVEY NOW INCLUDES income imputation, which marks a substantial improvement in the data. Nevertheless, researchers need to be aware of the consequences of income imputation. This article outlined four possible consequences. First, a measure of expenditure outlays is used to impute income, which means that there may be an artificial dependence between income and expenditures. Second, imputed income also directly affects the BLS official published measure of expenditures. After imputing wage and salary income, BLS assumes each consumer unit pays Social Security taxes from its wage or salary income. The amount of Social Security taxes is added to the official personal insurance and pensions category and in the official BLS measure of total expenditures.

Third, for data from 2004 forward, researchers will not have to restrict the CE sample to complete income reporters if they want to examine income and expenditures together. Because in-

complete income reporters are more likely to have lower consumption expenditures than complete income reporters, research that focuses on low consumption expenditure households is more likely to be affected by the use of all consumer units. The results in this study that were most affected by the inclusion of incomplete income reporters were the poverty rate

**Table 5. Ordinary least squares estimates of Engel curves for food expenditures, 2002–04**

Item	Complete income reporters	All consumer units
All consumers:		
Consumption expenditures .....	-0.329 (.020)	-0.340 (.020)
Consumption expenditures squared ..	.012 (.001)	.012 (.001)
Consumption-expenditure poor:		
Consumption expenditures .....	.630 (.246)	1.075 (.227)
Consumption expenditures squared ..	-.043 (.014)	-.067 (.013)

SOURCE: Consumer Expenditure Interview Survey (2002–2004). All data are population weighted. Expenditure and income data are in real 2004 dollars using the CPI-U-RS. Consumer units that have consumption-expenditures below the official Census poverty threshold are classified as consumption-expenditure poor. The poverty thresholds are the official thresholds published by the U.S. Census Bureau.

NOTE: The dependent variable equals the share of food expenditure in total consumption expenditures. The other independent variables in each regression are age, the number of adults, the number of children, and the square of each. There are also dummy variables for region, education, race, and year. Figures in parentheses represent standard errors.

and the Engel curve for consumption expenditure poor households. Alternatively, the Gini coefficient and the Engel curve for all households were not affected by the inclusion of incomplete reporters. Fourth, there may be a break in time series of statistics that use 2004 data in combination with earlier data. This article has shown that the change in the poverty rate between 2003 and 2004 depends greatly on what sample is used for each year. The poverty rate is much higher when using all consumer units than when using complete income reporters.

This analysis has only used the Interview portion of the CE, but it applies to the Diary survey as well. Both surveys now impute from 2004 forward, and both use the same imputation

approach. All of the consequences described for the Interview Survey also apply to the Diary Survey.

Overall, there are legitimate concerns about using the expenditure data along with imputed income data. There may be a temptation to continue using the complete income reporter sample rather than using all consumer units. As researchers become aware of the potential issues, they probably will want to use all consumer units starting in 2004. The main advantage to using all consumer units is in the precision of estimates. In 2004, 13 percent of all consumer units were classified as incomplete income reporters. By having a larger sample size after imputation, the precision of analysis using the CE will increase. □

## Notes

<sup>1</sup> See T. I. Garner and L. A. Blanciforti, "Household Income Reporting: An Analysis of U.S. Consumer Expenditure Survey Data," *Journal of Official Statistics*, March 1994, pp. 69–91.

<sup>2</sup> All results are population weighted. The data appendix describes the weighting methodology.

<sup>3</sup> Rubin, D.B., *Multiple Imputation for Nonresponse in Surveys* (New York, Wiley & Sons, 1987).

<sup>4</sup> Specifically, the quarterly expenditure–outlays variable is named *ERANKMTH*, and it equals the dollar amount of expenditure outlays made during the last 3 months.

<sup>5</sup> For a more detailed description of the BLS imputation procedure, see G. Paulin, S. Tsai, and M. Grance, "Model-Based Multiple Imputation," Paper 210–29, *Proceedings of the Twenty-Ninth Annual SAS Users Group International Conference* (Cary, NC, SAS Institute Inc., 2004).

<sup>6</sup> Officially, 6.2 percent is paid for Old–Age Survivors and Disability Insurance, and 1.45 percent is for Medicare.

<sup>7</sup> *Consumer Expenditures in 2004*, USDL 05-2243 (U.S. Department of Labor), November 29, 2005; available on the Internet at <http://www.bls.gov/news.release/cesan.nr0.htm>

<sup>8</sup> See, for example, J. M. Rogers and M. Gray, "CE Data: Quintiles of Income versus Quintiles of Outlays," *Monthly Labor Review*, December 1994, pp. 32–37. J. D. Fisher, D. S. Johnson, J. Marchand, T. Smeeding, and B. B. Torrey, "The Retirement Consumption Conundrum: Evidence from a Consumption Survey," CRR WP 2005–14, (Center for Retirement Research at Boston College, December 2005).

<sup>9</sup> Researchers will still have the option to restrict the sample to complete income reporters. The variable identifying which households are complete income reporters is included in the 2004 wave and the 2005 wave of the CE.

<sup>10</sup> G. Paulin, "A Changing Market: Expenditures by Hispanic

Consumers, Revisited," *Monthly Labor Review*, August 2003, pp. 12–35; S. Phipps and T. I. Garner, "Are Equivalence Scales the same for the United States and Canada?" *Review of Income and Wealth*, March 1994, pp. 1–17; O. Attanasio, E. Battistin, and A. Leicester, "Inequality, Poverty, and Their Evolution in the US: Consumption and Income Information in the Consumer Expenditure Survey," *Notes prepared for the National Poverty Center's ASPE-Initiated Workshop on Consumption among Low-income families*, 2004; available on the Internet at [http://www.npc.umich.edu/research/npc\\_research/consumption](http://www.npc.umich.edu/research/npc_research/consumption).

<sup>11</sup> Garner and Blanciforti find a similar pattern for the ages of complete and incomplete reporters in "Household Income Reporting."

<sup>12</sup> Using data from 1991–2004, R. Bavier also finds that mean expenditures are lower for incomplete income reporters. See R. Bavier, "Income and Expenditure Data in Poverty Measurement," unpublished manuscript, 2006.

<sup>13</sup> For an example of research that examines consumption inequality, see D. S. Johnson, T. M. Smeeding, and B. B. Torrey, "Economic inequality through the prisms of income and consumption," *Monthly Labor Review*, April 2005, pp. 11–24.

<sup>14</sup> See D. M. Cutler and L. Katz, "Rising Inequality? Changes in the Distribution of Income and Consumption in the 1980's," *American Economic Review*, May 1992, 546–51; D. T. Slesnick, "Gaining Ground: Poverty in the Postwar United States," *Journal of Political Economy*, February 1993, pp. 1–38; and M. Federman, T. I. Garner, K. Short, W. B. Cutter IV, J. Kiely, D. Levine, D. McGough, and M. McMillen, "What does it mean to be poor in America?" *Monthly Labor Review*, May 1996, pp. 3–17.

<sup>15</sup> These results are very similar to the ones presented in Attanasio and others, "Inequality, Poverty, and Their Evolution in the US."

<sup>16</sup> This again matches the findings of Attanasio and others, *Ibid.*

## APPENDIX: Data

---

The consumption expenditures measure includes the amount that the consumer unit actually spends for current consumption. This includes expenditures for food, housing, transportation, apparel, medical care, entertainment, gifts (of cash, goods and services) to organizations or persons outside the consumer unit, and miscellaneous items for the consumer unit. Excluded are expenditures for pensions and social security, savings, and life insurance.

Housing includes expenses associated with owning or renting a home or apartment, including rental payments, mortgage principal and interest, property taxes, maintenance, repairs, insurance, and utilities.

Transportation includes expenditures for the net purchase price of vehicles, finance charges, maintenance and repairs, insurance, rental, leases, licenses, gasoline and motor oil, and public transportation. Public transportation includes fares for mass transit, buses, airlines, taxis, school buses, and boats.

Medical care expenditures are for out-of-pocket expenses including payments for medical care insurance, medical services, and prescription drugs.

Entertainment expenditures are for fees and admissions, televisions, radios, sound equipment, pets, toys, playground equipment, and other entertainment supplies, equipment, and services.

Miscellaneous expenditures are for personal care services, reading, education, tobacco products and smoking supplies, alcoholic beverages, other lodging, and house furnishings and equipment.

All expenditure data in this article are adjusted using the Consumer Price Index research series (CPI-U-RS). The data are weighted using the weight variable *FINLWT21*. Because young renters are underrepresented in the sample of consumer units who remain in the survey for all five interviews, a procedure presented by J. Sabelhaus is used to adjust the weight variable by age and housing tenure (homeowner or renter) to obtain a better representation of the population as a whole.<sup>1</sup>

For measures of inequality such as the Gini coefficient, it is desirable to have a comparable unit of measurement. It is difficult to compare the consumption expenditures of a single-person consumer unit with those of a four-person consumer unit. The four-person consumer unit is expected to have higher consumption expenditures when everything else remains equal. To overcome this difficulty, the consumption expenditures of a consumer unit are adjusted using an equivalence scale, which allows for comparisons across consumer units of different sizes. The scale is given by the square root of the family size and indicates that the resources for a four-person consumer unit must be twice that of a single-person consumer in order for the two to have an equivalent standard of living. The equivalence scale is only used for calculation of the Gini coefficient.

## Note to the appendix

---

<sup>1</sup> J. Sabelhaus, "What is the Distributional Burden of Taxing Consumption?" *National Tax Journal*, September 1993, pp. 331–44.



