Published in final edited form as:

Am Econ Rev. 2014 May ; 104(5): 132–135. doi:10.1257/aer.104.5.132.

Estimates of Annual Consumption Expenditures and Its Major Components in the PSID in Comparison to the CE †

Patricia Andreski, Geng Li, Mehmet Zahid Samancioglu, and Robert Schoeni*

Andreski: Office for Health Equity and Inclusion, University of Michigan Health System, 1301 Catherine St., Ann Arbor, Michigan 48109 (pandresk@umich.edu); Li: Federal Reserve Board, 20th and C St., NW, Washington, DC 20551 (geng.li@frb.gov); Samancioglu: Department of Economics, University of Michigan, 611 Tappan St., Ann Arbor, Michigan 48109 (mzs@umich.edu); Schoeni: Institute of Social Research and Ford School of Public Policy, University of Michigan, 611 Tappan St., Ann Arbor, Michigan 48109

Historically, empirical research on consumption primarily relied on aggregate data, such as the National Income and Product Accounts (NIPA) tables. Beginning in the 1980s, studies of consumption expenditures using household survey data became more common. The majority of this research used the Consumer Expenditure Survey (CE), which for nearly two decades had been essentially the only national survey that provided extensive data on household consumption expenditures. Though rich in detailed expenditure data, the CE has more limited information on income, assets, and other socioeconomic factors. In addition, the CE is a short panel that follows each household for at most four quarters, limiting the extent to which these data can be used to study dynamic behaviors of households.

By contrast, the Panel Study of Income Dynamics (PSID) is a longitudinal survey that collects extensive information on socioeconomic characteristics, labor market experiences, income, wealth, health status, and family structure. However, historically only housing and food-related expenditure data were collected in the PSID.

In light of the growing interest in household consumption behaviors, the PSID expanded its questions on consumption expenditures significantly in 1999, and even further in 2005. As a result, the PSID now covers essentially all major expenditure categories that are included in the CE.

In this article, we describe the features of the PSID survey that make it a unique resource for studying household consumption. We then present statistical results that show that the PSID expenditure data compare favorably to their CE counterparts. We take the CE data as our reference point and a comparison benchmark because despite the fact that it has been demonstrated that estimates of aggregate consumption derived using the CE data fall short of estimates based on the NIPA data, the CE remains the most comprehensive household

†Go to <http://dx.doi.org/10.1257/aer.104.5.132> to visit the article page for additional materials and author disclosure statement(s).

*(bschoeni@umich.edu).

The views presented in this paper are those of the authors and are not necessarily those of the Federal Reserve Board or its staff.

expenditure survey. Moreover, Bee, Meyer, and Sullivan (2012) documented that most of the major expenditure categories in the CE are fairly consistent with the NIPA estimate.

I. The Two Surveys

A. The PSID

The PSID is the world's longest running national longitudinal household survey. Started in 1968, the PSID was an annual survey through 1997 and a biennial survey afterward. A hallmark feature of the PSID sample design is that it follows not only the original 1968 sample but also their descendants as they grow up and form their own households. In addition, the PSID has consistently achieved low sample attrition rates. As of the 2009 survey, the PSID collected data from nearly 8,700 households. Moreover, rich information on health status and behaviors, household balance sheets, time use, philanthropy, child development, and many other domains were added to the survey over time. In sum, the length of the panel, its genealogical design, and its broad content make it a unique data source for studying a wide array of topics on household behaviors. For example, the PSID data allow analyses of the transmission of economic behavior and well-being—such as consumption, saving, and wealth accumulation—across generations.

B. The CE

The CE is a nationally representative survey with a primary goal of collecting detailed data on household spending in order to inform estimates of the Consumer Price Index. The CE also collects key demographic and socioeconomic information of the surveyed household, making it a valuable source for economic research.

The CE has two components—an Interview Survey and a Diary Survey—that are administered to separate samples. The PSID is more comparable to the Interview Survey in terms of survey methodology because both surveys typically ask the consumers to recall their expenditures over the prior several months, in contrast to the bookkeeping approach of the Diary Survey, whose focus is on expenditure items that are frequently purchased, such as food. Accordingly, in this paper the PSID expenditure data are compared to data from the CE Interview Survey only. As of 2010, about 7,000 consumer units were interviewed each quarter. Compared with the PSID, the CE Interview Survey collects data at a higher frequency but maintains a much shorter longitudinal structure.

II. Expenditure Data in the PSID

The questions used in the PSID to collect consumption expenditure data and the years these questions were included in the survey are provided in online Appendix Table 1. Before 1999, the PSID collected only limited information on consumption expenditures. Partly because of its historical focus on poverty, a select set of poverty-related spending items, such as food and housing, was collected in almost all waves. Earlier research used this limited set of data as a proxy for consumption behavior of the household (e.g., Hall and Mishkin 1982; Hotz, Kydland, and Sedlacek 1988; Zeldes 1989). The food and rental expenditure data also served as inputs for estimating total expenditure by means of using estimated coefficients from the CE data (Skinner 1987).

To expand the coverage of expenditure data in PSID, a series of new questions was added to the survey in 1999 to collect information regarding spending on healthcare, education, childcare, transportation, and utilities. Li et al. (2010) find that the PSID-measured consumption expenditures in these categories match fairly closely to those measured in the CE and account for 72 percent of total consumption expenditures measured in the CE. The consumption expenditure questions were further expanded in 2005 to include information on spending on home repairs and maintenance, household furnishing, clothing, trips, vacations, and entertainment. With these additions, since 2005 the PSID has captured almost all expenditures measured in the CE.

In comparison to the CE, consumption data collected in the PSID have three distinct features that help to improve the chance of collecting valid responses, as indicated in online Appendix Table 2. First, for many categories of spending the PSID allows respondents to report expenditures in a time frame that is easiest for them to recall. Furthermore, although the survey questionnaire may ask the respondent to report over a specific time frame (e.g., “last month”), respondents can and do report for different time periods (e.g., “annual” or “weekly”). For example, as shown in the table, while the survey asks about spending on food at home in an average *week*, about 8 percent of the households in the survey prefer to report a monthly amount. Allowing for flexibility in reference time period and frequency likely decreases nonresponse rates and improves data quality.

Second, for some expenditure categories the PSID offers the respondents unfolding brackets when they cannot recall the exact amount of expenditures. For example, if in 2009 a respondent reported “don't know” for health insurance expenditures, they were then asked “were they \$2,500 or more?” If the respondent says yes (no), the follow-up question asked whether they spent “\$5,000 or more (\$1,000 or more)?” This approach further reduces item nonresponse and strengthens data quality.

Third, unlike the CE, the PSID collects consumption expenditure information at a more aggregate level. The PSID is a general purpose survey that collects information on many aspects of household activity. As a result, the time it can devote to collecting consumption expenditures data is more limited relative to the CE. As of the most recent wave of the PSID survey, only 48 questions were asked to collect consumption data, a small fraction of those in the CE. Consequently, the PSID collects less detailed information than the CE. For example, the PSID collects information on home repairs and maintenance by asking “How much did you spend altogether on home repairs and maintenance, including materials plus any costs for hiring a professional?” This same information was reflected in more than 50 Universal Classification Code (UCC) items in the CE data, each of which is collected separately.

III. Comparing the PSID and the CE Consumption Expenditure Data

Three steps are taken in order to compare the expenditure data in the PSID and the CE. First, the detailed CE expenditure items have to be mapped into the broader expenditure categories of the PSID. To do so, we build on the mapping in Li et al. (2010), establishing a new mapping for the PSID expenditure categories introduced in 2005. As in Li et al. (2010), the

mapping was judgmental and subject to multiple rounds of independent review. Online Appendix Table 3 reports the specific CE UCC codes that are mapped to each PSID expenditure question.¹ As noted in this table, some UCC codes are not captured by any PSID consumption expenditure question, and this accounts for approximately 5 percent of CE spending.

Second, because survey respondents in the PSID are given the option of answering several expenditure questions using unfolding brackets, such categorical responses need to be converted into exact amounts. We estimate the conditional mean expenditure for each bracket using the exact-number-responses that fall into this particular bracket.² These mean estimates are assigned to the households who responded using the option of unfolding brackets. For example, a household who responded to the survey that its health insurance expenditure was between \$1,000 and \$2,500 will be assigned the mean estimates of all households whose reported health insurance expenditure was between \$1,000 and \$2,500.

The third step is to impute values for respondents who did not provide a valid response to a given expenditure question. Although response rates for each specific question in the PSID are high for almost all categories of spending (online Appendix Table 2), the implied response rate for total expenditure (the sum of spending on all expenditure categories) can be significantly lower. For example, 98–99 percent of respondents provide valid responses to most questions about specific categories of transportation spending, but just 85 percent of responding households have valid responses to all components of transportation expenditures. Imputation is conducted by relating valid responses of each expenditure category to family size and age of household head. Specifically, we fit an equation with a vector of family size dummies and a cubic polynomial of household head age.

Online Appendix Table 4 presents PSID-based expenditure estimates for each year since PSID expanded its consumption expenditure questions (i.e., 1999, 2001, 2003, 2005, 2007, and 2009 waves), as well as comparisons with their counterparts in the CE data. Note that all dollars are reported in nominal values to facilitate comparison with published tabulations by the CE. The expenditure data in the two surveys line up well for the categories of spending collected by the PSID; the ratio of mean PSID spending to mean CE spending varied within a fairly narrow range between 0.96 and 1.02 in survey years 1999 through 2009, respectively. Furthermore, as shown in online Appendix Figure 1, the life-cycle expenditure profiles estimated using the PSID and the CE data are very similar, and the CE profile stays mostly within the 95 percent confidence interval of the PSID profile.

That said, larger gaps between the PSID and CE are evident in certain expenditure categories and subcategories. For example, the PSID education expenditure was 86 percent of the CE estimate in 2009. Expenditures on repairs and maintenance of vehicles also differ substantially between the two data sources and this gap varies considerably over time. In addition, the new expenditure categories introduced to the PSID in 2005 appear to

¹This table reports the mapping for the latest wave used for the comparison. Mapping is updated for each year to reflect changes in UCC.

²An alternative approach is to use the median value for each bracket except the top bracket. The two approaches yield results that are qualitatively very similar.

demonstrate much larger differences relative to the CE mean estimates than for categories of spending that were assessed in earlier waves. Home repairs and maintenance in the PSID were roughly twice as large as the CE estimates, and the PSID clothing and apparel expenditure is also substantially higher than the CE. Even for the expenditure categories with mean estimates reasonably close between the two surveys, statistics for subcategories can be very different. For example, while total housing spending is only 2–4 percent higher in the PSID, the PSID home insurance expenditures are 40–50 percent higher.

These differences most likely arise because of differences in both the sampling and survey methodology in the two surveys. In particular, we suspect that for some expenditure components, it might be easier for consumers to recall the total spending at a high level of aggregation, whereas for other expenditures, the recall is more accurate at a more detailed level. In addition, although the current mapping between UCCs and the PSID consumption expenditure categories represents our best effort, there might be some UCCs that are miscategorized, leading to greater gaps at more detailed expenditure categories but not affecting total expenditure estimates. Indeed, although the PSID and CE's respective estimates for home repairs and maintenance and household furnishings and equipment are off by a large margin, the sum of the two estimates compare more favorably between the two surveys. Furthermore, respondents might report spending in the incorrect category. For example, some respondents may report some types of household equipment expenditure as a home repair. Similarly, some types of recreation and entertainment may be misreported as trips and vacations, perhaps explaining the fact that the PSID based-estimate of the former is lower than CE while the PSID based-estimate of the latter is higher than CE.

Finally, we examine whether consumption expenditures measured in the PSID and the CE vary in similar ways with observable household characteristics. Specifically, we regress log total expenditure on log family total income, a cubic polynomial in age of the head, vectors of household size and educational attainment, and dummies of marital status, race, and homeownership. We also controlled for yearly fixed effects. The results (not shown) coefficients are broadly consistent between the suggest that the estimated PSID and the CE samples.

IV. Concluding Remarks

A decade-long effort to boost consumption expenditure data in the PSID gives the research community a new data source for examining household consumption behavior. We document that PSID expenditure data are largely consistent with the CE data, in particular for total household expenditure. However, significant differences exist for some subcategories of expenditure. In addition, readers should be mindful that the CE has likely underestimated total expenditure relative to the NIPA data. Indeed, the CE is undergoing a major overhaul of its own survey design to improve the data consistency and accuracy, which warrants future periodical comparison between the two surveys.

With these caveats in mind, taking advantage of the multi-generational, long panel structure of the survey, and the extensive data on labor market outcomes, health status and behaviors,

sociodemographic factors, and balance sheets, a wide array of topics related to household consumption now can be studied using data representative of the full US population.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We acknowledge Richard Blundell and James Sullivan for helpful discussions and comments, Ben Gage Love for able research assistance, and grants from NIA (AG019802) and NSF (SES0518943).

REFERENCES

- Bee, Adam, Meyer, Bruce D., Sullivan, James X. National Bureau of Economic Research Working Paper 18308. 2012. The Validity of Consumption Data: Are the Consumer Expenditure Interview and Diary Surveys Informative?.
- Hall, Robert E., Mishkin, Frederic S. The Sensitivity of Consumption to Transitory Income: Estimates from Panel Data on Households. *Econometrica*. 1982; 50(2):461–81.
- Hotz, V. Joseph, Kydland, Finn E., Sedlacek, Guilherme L. Intertemporal Preferences and Labor Supply. *Econometrica*. 1988; 56(2):335–60.
- Li, Geng, Schoeni, Robert F., Danziger, Sheldon, Charles, Kerwin Kofi. New Expenditure Data in the PSID: Comparisons with the CE. *Monthly Labor Review*. 2010; 133(2):29–39.
- Skinner, Jonathan. A Superior Measure of Consumption from the Panel Study of Income Dynamics. *Economics Letters*. 1987; 23(2):213–16.
- Zeldes, Stephen P. Consumption and Liquidity Constraints: An Empirical Investigation. *Journal of Political Economy*. 1989; 97(2):305–46.