

Consumer Expenditure Survey Microdata Users' Workshop, July 2010

Geoffrey Paulin

The Consumer Expenditure Survey (CE) is the most detailed source of expenditures, demographics, and income data collected by the Federal Government. The data are collected in two component surveys: the Quarterly Interview Survey (henceforth referred to as the Interview Survey) and the Diary Survey. Each year, the CE program releases microdata from these surveys; these microdata are used by researchers in a variety of areas, including academia, government, market research, and other private industry.

Since 2006, the Division of Consumer Expenditure Surveys (DCES) has conducted an annual workshop each July for users of the CE microdata. Held in the conference facilities of the Bureau of Labor Statistics (BLS) headquarters in Washington, D.C., the workshops have included speakers demonstrating features of the data, as well as reports from researchers who have used these data in their work. Each year, the format has changed to incorporate suggestions from participants, but the basic framework has remained intact.

In July 2009, the program was expanded from two days to three days. The first day was designed especially for new users, including

novices and those who had never used the data. The second day was designed to feature research from users outside the BLS. The third day was designed especially for more experienced users. The program was arranged in this way to accommodate as many potential participants as possible. That is, any attendee could participate in one, two, or all three days of the workshop and benefit from sessions geared toward his or her expertise.

July 2010 workshop

The July 2010 workshop featured a slightly different format. Because of comments from the 2009 workshop, research presentations were spread out over the three days. Nevertheless, the training and data-descriptive sessions were organized progressively so that participants could attend the combination of days appropriate to their levels of expertise in using the data. The speakers at the workshop did an excellent job presenting not just results of their work, but processes used, problems or data limitations encountered and how they were handled, and other practical considerations.

Finally, a new feature called “meet with an expert” was initiated. In this feature, participants had the opportunity to make one-on-one appointments with an expert data user from the staff of the Consumer Expenditure Survey program for an in-depth discussion about their specific or general questions regarding the data or its uses. Several participants did so.

First day. The first day of the 2010 workshop opened with an overview of the CE, featuring topics such as how

the data are collected and published (Veri Crain). The overview was followed by a research presentation that combined data from the CE and the Consumer Price Index to discover whether changes in spending patterns at regional levels were due to price changes, population changes, or other factors (Cassandra Wirth, Midwest BLS Information Office in Chicago). Next was an introduction specifically to the microdata, including an explanation of its features (Bill Passero and Jeff Crilley), which was followed by a research presentation on estimation of wives' work-related costs in dual-earner households (Seonglim Lee, Sungkyunkwan University, South Korea). The afternoon included research presentations on expenditures for frozen and prepared meals (Megumi Omori, Bloomsburg University) and comparisons of charitable contributions by men and women (Sanae Tashiro, Rhode Island College). These presentations were followed by two practical hands-on training sessions with expert users from the DCES staff (Laura Paszkiewicz and Crilley); the presenters demonstrated appropriate use of the files and variables to obtain estimates while participants practiced together on shared laptops.¹

Second day. The second day began with presentations on advanced topics, including technical details about sampling methods and construction of sample weights (Catherine Hackett, Division of Price Statistical Methods), imputation and allocation of microdata (Troy Olson), and common questions

Geoffrey Paulin is a senior economist in the Division of Consumer Expenditure Surveys, Bureau of Labor Statistics. Email: paulin.geoffrey@bls.gov

about “calendar” versus “collection” period expenditures (Passero).² Following this, a practical training session described specific steps required to compute calendar year estimates, both unweighted and weighted (Paszkievicz and Crilley).

After a break for lunch, the afternoon opened with two research presentations. The first described expenditure patterns for low-income households, and estimated how eligibility for certain types of public assistance would differ under expenditure-based and income-based measures of poverty (James Mabli, Mathematical Policy Research, Inc.). This was followed by a presentation demonstrating specific programming code used in the computation of health care expenditures by Medicare households (Anthony Damico, The Kaiser Family Foundation). This presentation led naturally to the next practical training session, which covered procedures for merging data sets and then manipulating the results to compute statistical measures (Passero).

The day concluded with two presentations by BLS staff; the first was by Brian Baker and Casey Homan, editors of the *Monthly Labor Review*. Their presentation described the publication process from submission to printing for authors who were interested in having their works appear in this journal. Next came a brief “sneak peek” at changes to the microdata files that would occur with the release of the 2009 microdata, including a description of never-before-released “paradata” regarding the interview process itself, such as contact history and whether the interview was by personal visit or telephone (Steve Henderson).

Third day. The third day featured advanced topics: using data from participants in all four published inter-

views, rather than treating observations from each quarter independently (Passero); an explanation of how sales taxes are applied to expenditure reports during the data production process (Meaghan Smith, formerly Duetsch); proper use of imputed income data (Geoffrey Paulin); and proper use of sample weights in computing population estimates (Paulin). Specifically, the sample weights session noted that proper use of weights requires a special technique to account for sample design effects. If this technique is not employed, there will be incorrect estimates of variances and regression parameters.³ This session was followed by a research presentation on joint determination of life and health insurance (Ashish Kumar, State University of New York, Buffalo). The morning concluded with the workshop’s final practical training session, which featured discussion of a program included with the microdata for use in computing proper standard errors for means and regression results when using various kinds of data; unweighted non-income data; population-weighted non-income data; and multiply-imputed income data, both unweighted and population-weighted.

The afternoon started with two research presentations. The first described the use of the CE data as part of a transdisciplinary project studying obesity (Amanda Goldstein, Center for Rural Studies, University of Vermont). The second described work in progress by a researcher (Jeff Lundy, University of California, San Diego) who had recently been granted access to confidential data sets as part of the BLS “onsite researcher” program (<http://www.bls.gov/bls/blsresda.htm>). The research specifically investigates characteristics of consumers whose spending exceeds income, and analyzes them from a sociological

perspective. The day concluded with CE program staff soliciting feedback from the participants.

Future workshops

The next workshop will be held July 27–29, 2011. It will be free of charge to all participants, although advance registration is required. For more information about the 2010 and 2011 workshops, visit the CE website (<http://www.bls.gov/cex>) and look for “Annual Workshop” under the left navigation bar entitled “PUBLIC USE MICRODATA.” For direct access to this information, the link is <http://www.bls.gov/cex/csxannualworkshop.htm>.

Abstracts of 2010 presentations

Following are abstracts of the papers read at the 2010 conference, listed in the order in which the papers were presented, and based on summaries written by their authors:

Household spending patterns by region. Cassandra Wirth (formerly Yocum), economist, Office of Field Operations, Midwest BLS Information Office in Chicago,⁴ Division of Economic Analysis and Information, presented “Household Spending Patterns: A Comparison of Four Census Regions.” This paper reviews expenditures made by households in selected areas within four regions of the United States as defined by the U.S. Census Bureau. Following a model previously published in the *Monthly Labor Review*, the author breaks down these aggregate expenditures into five categories of change that can impact total expenditures: population growth within a geographic region, changes in population concentrations among local areas within a geographic region, changes in the definitions

of goods and services as collected and priced by the Consumer Price Index (CPI), price changes, and quantity changes. For each region, the paper discusses some of the largest impacts of each component. The paper provides comparisons of results among regions for selected goods and services in each of the eight major groups of commodities and services used by the CPI.⁵

Earnings contributions of wives. Seonglim Lee, associate professor, Department of Consumer and Family Sciences, Sungkyunkwan University (Seoul, South Korea), presented “The estimation of wife’s work-related costs in dual earner households.” This paper evaluates wives’ earnings contributions to household income in dual-earner households and to household income class and mobility. This work-in-progress uses data from the 2002–03 Interview Surveys to compare expenditure patterns for selected goods and services when comparing husband/wife consumer units where the wife is working full-time with consumer units where the wife is not employed outside of the home.

The change in expenditures on frozen and prepared foods. Megumi Omori, assistant professor of sociology, Department of Sociology, Social Work, and Criminal Justice, Bloomsburg University, presented “Expenditure of Frozen and Prepared Meals: 1980–2008 Consumer Expenditure Diary Data.” According to the literature, as women’s labor force participation has increased, hours spent on household chores have decreased in the U.S. over the past three decades. The reduction in household chore hours is often attributed to an increase in the frequency of dining out. However, over the past several years, there is little evidence of an increase

in dining out: the mean frequency of weekly dining out was approximately 1.3 in 1990 and 1.25 in 2006. One way to reduce meal-related household chore hours, aside from dining out, is to use prepared and frozen foods. Very little research has been done to examine a possible change in the use of prepared and frozen food. By using the Consumer Expenditure Diary Survey, the study in progress tries to find expenditures of frozen and prepared meals since 1980. Specifically, the study uses Universal Classification Codes (UCCs) 180210 (frozen meals), 180220 (frozen prepared food), and 180710 (miscellaneous prepared foods). Because these codes have remained the same since 1980, the study is able to directly compare consumer unit expenditures on these items over a nearly three-decade span. Although the presentation included results of the research, the primary focus of the presentation was the methods and applications used in studying the data.

Few gender differences in philanthropic giving. Sanae Tashiro, assistant professor of economics, Rhode Island College, presented “Are Women More Generous Than Men? Evidence from the U.S. Consumer Expenditure Survey.” Using data from the 2006 CE, the paper tested hypotheses based on theories of human and social capital by examining whether gender, age, education, income, race, and ethnicity affect giving. As a result of comments from the workshop, the paper has been revised to include Heckman’s two-stage sample selection estimates, which show that gender differences in philanthropic behavior are non-existent. Education, annual income, wealth, and Hispanic ethnicity increase the probability of giving but have no effect on the actual donation amount. Estimates further show that

age and race interact with gender to affect differences in giving—older women are more likely than younger men to donate but give smaller shares of income, while White women, Black women, and Asian women are less likely to donate, and those who do give smaller dollar amounts than do women of other races.

Expenditures of low-income households. James Mabl of Mathematica Policy Research, Inc. presented “Low-Income Household Spending Patterns and Measures of Poverty,” which was coauthored by Laura Castner, project director. The presentation described expenditure patterns for low-income households, how eligibility for certain types of public assistance would be expected to differ under expenditure-based and income-based measures of poverty, and how the estimates were obtained using the CE Interview Survey. In their report, the authors examined how low-income households in 2005 allocated income across consumption categories. The authors compared expenditures of participants in the Supplemental Nutrition Assistance Program (SNAP)—the Federal transfer program formerly known as the Food Stamp Program—with data for two other groups of low-income households: those whose income made them eligible for SNAP but did not participate, and households whose income exceeded SNAP eligibility limits. For each of the three groups, the authors estimated how a small increase in income would be allocated across each consumption category, and analyzed how eligibility for SNAP could change if it were based on expenditures rather than income. In addition, the authors explored the use of savings and credit across the three groups.⁶

Purchases by Medicare recipients. An-

thony Damico, The Kaiser Family Foundation, presented “Health Care on a Budget: An Analysis of Spending by Medicare Households.” This presentation was designed to teach users how to define any population of interest from among the interview files and then rapidly produce graphs and charts about any expenditure category of interest. First, the presenter explained how to narrow expenditure categories to only the ones of interest. A researcher might be interested in apparel and services, alcoholic beverages, or education; although the presentation used health-care as an example, a few nominal changes allow analysis of other categories of interest. Second, by making some minor edits to the SAS program included with the data set in the SAS programs folder—“Intrvw Mean and SE.sas”—one can limit the output to only the expenditures of interest, and the output can be broken out by any demographic group that one can identify by the family files. Third, in order to increase the number of ways to identify demographic groups, this presentation reviewed how to merge the family files with some of the other interview files. After completing those three steps, the researcher will have an output file containing the expenditure categories that he or she is most interested in, broken down and filtered according to precise analytic needs. Again, the example used was household health-care expenditure categories among Medicare beneficiaries, broken down by various demographic groupings. Finally, the presentation included a technique that can be used to quickly create an “all other” expenditure category, a category which combines multiple categories.

Determinants for choosing health and life insurance. Ashish Kumar, Ph.D.

candidate in marketing, School of Management, The State University of New York (SUNY) at Buffalo, presented “On the Relationship Between Health Insurance and Life Insurance Choice: A Disaggregate Level Analysis.” This paper investigated the joint determination of household choice for health insurance and life insurance. Using the 2008 Consumer Expenditure Survey data and assuming households consider purchasing health insurance and life insurance in order to manage life’s financial risks, the paper modeled household choice for those purchases after accounting for household characteristics, health and disability status, and insurance characteristics. The model that was used helped to assess the impact health insurance choice has on the choice of life insurance, and analyzed the correlation between these two choices. The result suggests that health insurance choice positively affects the choice of life insurance and that these two choices are positively correlated, which indicates that these two types of insurances are complementary.

A transdisciplinary approach to understanding obesity. Amanda Goldstein, MS., research associate, Center for Rural Studies, University of Vermont, presented “Comparison of Discipline Specific Food Categorization Within the Consumer Expenditure Survey When Examining Overweight.” Although obesity continues to be a public health concern in the United States and throughout the world, obtaining and maintaining a healthy weight is a decision—involving time and goods devoted to meal production and energy expenditure—that can only be made at the household level. Researchers tend to focus on obesity from distinct disciplinary perspectives rather than recognizing

that the obesity epidemic is multidimensional and that a transdisciplinary approach is required to gain a complete understanding of the obesity epidemic. Transdisciplinary work should be a flowing and adaptive process driven by collaboration in the design, implementation, and application of research. Current efforts suffer from lack of both shared language and terms of understanding across disciplines. The study investigates the magnitude of the differences in definition between economists and nutritionists and how such differences in definition impact the analyses of the production of a healthy weight. Each profession classified detailed food expenditure data from the Consumer Expenditure Survey into broad food groups. Average expenditures for each broad food group then were compared to test for discipline-based differences. Statistically significant differences in categorizations were found between the disciplines, as were sharply contrasting conclusions concerning what contributes to obesity. The study concludes that the epidemic is likely to continue to plague this country until a transdisciplinary approach to the problem integrates both across disciplines and across institutions.

Investigating why Americans overspend. Jeff D. Lundy, Ph.D. candidate in sociology, University of California, San Diego, presented “Keeping Up Appearances or Just Keeping Afloat: How and Why American Households Overspend?” To address why Americans overspend, this paper explores how overspending is distributed among American households, and compares the empirical conformity of that distribution with the expectations of prominent theorists. By examining which households buy which kinds of goods, this research advances

our understanding of the social and economic factors that contribute to overspending. The phenomenon is found to be widespread, but its extent varies depending on the demographic group and time period over which it is examined. Results suggest that over-

spending has the highest prevalence among low-income, non-wealth-owning households. Routine overspending is found to have little explanatory power. However, evidence suggests that indicators of a liquidity crisis are predictive of overspending.

The presentation also described the onsite researcher program, in which selected applicants can obtain access to confidential BLS microdata files to conduct approved statistical analyses (<http://www.bls.gov/bls/blsresda.htm>). □

Speakers at the workshop

BLS staff of the Division of Consumer Expenditure Surveys:

- Crain, Veri, economist, Branch of Information and Analysis; day 1
- Crilley, Jeffrey, economist, formerly with Branch of Information and Analysis; days 1 and 2
- Smith (formerly Duetsch), Meaghan, supervisory economist, Chief, Phase 1/Phase 2 Section, Branch of Production and Control; day 3
- Henderson, Steve, supervisory economist, Chief, Branch of Information and Analysis; days 1 and 2
- Olson, Troy, supervisory economist, Chief, Phase 3 Section, Branch of Production and Control; day 2
- Passero, Bill, senior economist, Branch of Information and Analysis; all days
- Paszkievicz, Laura, senior economist, Branch of Information and Analysis; days 1 and 2
- Paulin, Geoffrey, senior economist, Branch of Information and Analysis; day 3

Other BLS speakers:

- Baker, Brian, technical writer-editor, Office of Publications and Special Studies, *Monthly Labor Review*; day 2
- Hackett, Catherine, mathematical statistician, Division of Price Statistical Methods; day 2
- Homan, Casey, technical writer-editor, Office of Publications and Special Studies, *Monthly Labor Review*; day 2
- Wirth (formerly Yocum), Cassandra, economist, Office of Field Operations, Midwest BLS Information Office in Chicago,⁷ Division of Economic Analysis and Information; day 1

Speakers from outside BLS:

- Damico, Anthony, The Kaiser Family Foundation, "Health Care

- on a Budget: An Analysis of Spending by Medicare Households" (Interview Survey), day 2
- Goldstein, Amanda, MS., research associate, Center for Rural Studies, University of Vermont, "Comparison of Discipline Specific Food Categorization Within the Consumer Expenditure Survey When Examining Overweight" (Diary Survey), day 3
- Kumar, Ashish, Ph.D. candidate in marketing, The State University of New York (SUNY) at Buffalo, "Joint Determination of Health Insurance and Life Insurance Choice Using Data from the Consumer Expenditure Survey" (Interview Survey), day 3
- Lee, Seonglim, associate professor, Department of Consumer and Family Sciences, Sungkyunkwan University, Seoul, South Korea, "The Estimation of Wife's Work-Related Costs in Dual Earner Households" (Interview Survey), day 1
- Lundy, Jeff, Ph.D. candidate in sociology, University of California, San Diego, "Accessing Confidential CE Microdata—with an example of research using the confidential data set" (Interview Survey), day 3
- Mabli, James, Mathematica Policy Research, Inc., "Low-Income Household Spending Patterns and Measures of Poverty," co-authored with Laura Castner, project director (Interview Survey), day 2
- Omori, Megumi, assistant professor of sociology, Department of Sociology, Social Work, and Criminal Justice, Bloomsburg University, "Expenditure of Frozen and Prepared Meals: 1980–2008" (Diary Survey), day 1
- Tashiro, Sanae, assistant professor of economics, Rhode Island College, "Are Women More Generous than Men?: Evidence from the U.S. Consumer Expenditure Survey Data" (Interview Survey), day 1

Notes

¹ Topics covered included a brief overview of the microdata files and structure, summary variables (i.e., aggregated values for various expenditure categories), and estimating unweighted and weighted mean expenditures using FMLY, MEMB, and MTAB files from the Interview Survey, and FMLY and EXPN files from the Diary Survey. For each survey, the FMLY file contains information for the consumer unit as a whole, such as region of residence and summary variables for expenditure categories such as total expenditures, housing, and apparel in the Interview Survey; and fresh fruits, fresh vegetables, nonalcoholic beverages, and nonprescription drugs and supplies in the Diary Survey. (For the definition of a consumer unit, see “About the CE Data.”) The MEMB files contain information for each member of the consumer unit, such as age, ethnicity, and educational attainment. However, the files contain no expenditures, as expenditure data are collected for the consumer unit as a whole, and therefore are not available for specific members, except in single-member consumer units. The MTAB files in the Interview Survey include information on expenditures at very detailed levels (e.g., food or board at school; rent of dwelling; bedroom linens; girls’ hosiery; and boys’ footwear). The

EXPN files in the Diary Survey are similar to the MTAB files in the Interview Survey, in that they also include information on expenditures at very detailed levels (e.g., apples; bananas; oranges; other fresh fruits; and citrus fruits excluding oranges, which aggregate to form the summary variable “FRSHFRUT” in the Diary Survey FMLY file). There are also files called EXPN in the Interview Survey, which contain even more detailed breakdowns for certain expenditures, and other detailed information for some items, such as the number of members of the consumer unit that are covered by a particular health insurance policy. However, the EXPN files from the Interview Survey were not discussed in this part of the training.

² In the Interview Survey, the three-month recall period may include expenditures made in a prior year. For example, persons interviewed in February will report expenditures occurring in November and December of the prior year, as well as expenditures occurring in January of the current year. Those interested in computing expenditures for the collection period can sum expenditures for these three months to obtain their results. However, those interested in computing expenditures that occurred within the same calendar year must

take extra steps to include the November and December expenditures only with prior year expenditures, and those made in January only with current year expenditures.

³ The CE sample design is pseudo-random. However, proper use of weights requires use of the method of Balanced Repeated Replication.

⁴ At the time the presentation was submitted to the workshop planning team for consideration, Ms. Wirth was employed by the Mountain-Plains BLS Information Office in Kansas City. She later moved to the Midwest BLS Information Office in Chicago.

⁵ This abstract is based on an abstract which can be found at <http://www.bls.gov/osmr/abstract/ec/ec070110.htm> (visited November 15, 2010). The full working paper can be found at <http://www.bls.gov/osmr/pdf/ec070110.pdf> (visited April 26, 2011).

⁶ This abstract is based on a report which can be found at <http://www.fns.usda.gov/ora/menu/Published/snap/FILES/Participation/SpendingPatterns.pdf>, Executive Summary, p. xiii (visited November 16, 2010).

⁷ See note 4.

APPENDIX: About the CE data

Consumer unit. The basic unit of analysis in the Consumer Expenditure Survey (CE) is the consumer unit. In general, a consumer unit consists of (1) all members of a particular household who are related by blood, marriage, adoption, or some other legal arrangement; (2) a person living alone or sharing a household with others or living as a roomer in a private home or lodging house or in permanent living quarters in a hotel or motel, but who is financially independent; or (3) two or more persons living together who use their incomes to make joint expenditure decisions. Financial independence is determined by spending behavior with regard to the three major expense categories: housing, food, and other living expenses. To be considered financially independent, the respondent must provide at least two of these expenditure categories, either entirely or in part.

Collection and methodology. Since 1980, the Interview and Diary Surveys have been collected on an ongoing basis. The Inter-

view Survey is designed to collect expenditures for big-ticket items (for example, major appliances, and cars and trucks) and recurring items (for instance, payments for rent, mortgage, and insurance). Data on some expenditures, such as food at home, are collected globally.¹ In addition to data on expenditures, demographics, and income, information about assets and liabilities is collected. In this survey, participants are visited once every 3 months for five consecutive quarters. Data from the first interview are collected only for bounding purposes and are not published.² Since April 2006, about 7,000 consumer units have participated each quarter.

In the Diary Survey, participants record expenditures daily for two consecutive weeks. The survey is designed to collect expenditures for small-ticket and frequently purchased items, such as detailed types of food (white bread, ground beef, butter, lettuce). Since April 2006, about 7,000 consumer units have participated annually. Because they complete a separate diary each week, approximately 14,000 diaries are collected each year.

Notes to the appendix

¹ That is, the respondent is asked to provide an estimate of the consumer unit's total expenditure for these items, rather than collecting detailed information on the items composing food expenditures.

² A bounding interview collects information to alert the interviewer to probe in cases where the purchase of a big-ticket or an infrequently purchased item reported in one interview is reported, perhaps inadvertently, in

the next interview. For example, if, in both the first and second interviews, the respondent reports that he or she purchased a refrigerator, the interviewer can ask followup questions during the second interview to ascertain whether the refrigerator that was purchased was the one reported in the first interview. The same process is followed in the second through fifth interviews when similar cases occur. That is, the second interview provides bounding information for the third interview, and so forth.