Consumer Expenditure Survey Microdata Users’ Workshop, July 2011

Geoffrey Paulin

The Consumer Expenditure Survey (CE) is the most detailed source of expenditures, demographics, and income collected by the federal government. CE data are collected by the U.S. Census Bureau in two component surveys: the Quarterly Interview Survey and the Diary Survey, which are described subsequently. (See Appendix A.) Every year, the CE program releases microdata from each of these surveys, which are used by researchers in a variety of areas, including academia, government, market research, and other private industry areas. In July 2006, the CE program office conducted the first in a series of annual workshops, to help users to better understand the structure of CE microdata; provide training in the uses of the surveys; and, through presentations by current users and interactive forums, promote conversation among the users, both to expand awareness within the community of common research interests and to discuss possibilities for collaboration. As part of these efforts, the past three workshops have also featured presentations by economists from the BLS regional offices who work with CE data. (See Appendix B.) The focus of this report is the most recent workshop, which was held in 2011.

Day one. William Hawk opened the first day of the 2011 workshop with an overview of the CE in general, featuring topics such as how the data are collected and published. Bill Passero and Craig Kreisler then presented an introduction to the microdata, including an explanation of its features, such as “topcoding.”

The session that followed included a new feature for the workshop: presentations by first-time users of the CE microdata. The first presenter, Karen Ransom of the Southeast BLS Information Office in Atlanta, spoke about research into the effects of the recent “housing bubble” in various major metropolitan areas across the United States. The second presenter, Kara Markley of the Mid-Atlantic BLS Information Office in Philadelphia, described her research into costs of commuting for residents of the Washington, DC, area, with particular emphasis on transit subsidies received by many of these commuters. Both presenters received their first training in use of the microdata within a few months of the workshop: Markley in March 2011 and Ransom in May 2011. The idea behind their presentations was to share their experiences with other novice users, including any tips or advice they had regarding working with the data.

The final speaker of the morning session, Tian Luo, of the Western BLS Information Office, a more experienced user of CE microdata, presented research on expenditures for higher education by the race of the reference person of the consumer unit. Like the others in this session, Luo is part of the Economic Analysis and Information (EA&I) staff at a BLS regional office. Since the 2009 workshop, at least one speaker from an EA&I staff has been included in the program. The inclusion of such a speaker is important to researchers, because EA&I staff can provide information on data relevant to their local areas not only from the CE program, but from other programs sponsored by BLS (e.g., data on unemployment rates in Chicago). As resources and circumstances permit, future workshops will be planned to include both additional sessions aimed at highlighting experiences of first-time CE microdata users and continued presentations by EA&I staff of BLS regional offices.

Jonathan Fisher, of the Census Bureau, started the afternoon session with a presentation on how households adjust to changes in gasoline prices. The presentation was interesting for many reasons, including the fact that it used microdata from both the Diary and Interview Surveys. Jeff Lundy, Ph.D. candidate in sociology at the University of California at San Diego, then spoke about annual changes in wealth as measured by the Interview Survey.

The afternoon concluded with the first of two practical “hands-on” training sessions, with Laura Paszkiewicz and Craig Kreisler, expert users from the CE program staff, demonstrating how to use the files and variables to obtain estimates. During the session, participants practiced together on laptop computers.

Day two. The second day opened with advanced topics, with Catherine Hackett of the BLS Division of Price Statistical Methods presenting technical details about sampling methods and construction of sample weights; Troy Olson of the CE program speaking on imputation and the allocation of microdata questions; and Bill Passero talking about common “calendar” versus “collection” period expendi-
Following these presentations, Laura Paszkiewicz and Craig Kreisler held a practical training session and described specific steps required to compute calendar-year estimates, both unweighted and weighted.

After a break for lunch, the afternoon opened with an informational presentation followed by two research presentations. First, Jeff Lundy described his experiences in the BLS onsite researcher program, in which approved researchers can obtain access to confidential data provided by the Bureau. Next, Adam Bee, Ph.D, Notre Dame University and soon-to-be Census Bureau employee, described his research into the relationship between car ownership and employment. Finally, Megumi Omori, of Bloomsburg University, described her research into gift expenditures during the holiday season, a study that used data from the Diary Survey. These presentations were followed by Bill Passero's practical training session, which covered procedures for merging datasets and then manipulating the results to compute statistical measures.

The day concluded with special concurrent sessions. First, Terry Schau, managing editor, and Brian Baker, technical writer-editor, from the Monthly Labor Review described the publication process, from submission to printing, for authors interested in having their works appear in that journal. Concurrently, a panel from the Committee on National Statistics, established to make recommendations for a redesign of the CE data collection instrument, spoke on how the members consulted with expert users in order to understand how researchers use CE data and to solicit their ideas on how the data might be improved. Next, Steve Henderson delivered a brief “sneak peek” at changes to the microdata files that were scheduled to occur with the release of the 2010 microdata in September 2011. Finally, participants in an information-sharing group discussed each other’s work with the data, their experiences at the workshop, and other topics in an unstructured setting.

Day three. The final day featured advanced topics, starting with Bill Passero's presentation of the use of data from survey respondents who completed all four published interviews. The issue was whether to combine the data collected or, instead, treat observations from each quarter independently. Coincidentally, Jonathan Fisher had developed his own weighting scheme for handling four-interview participants, and he described the methodology in the same session, marking the first time in the history of these workshops that circumstances combined to allow the sharing of a presentation between a BLS and a non-BLS researcher. In subsequent presentations, Neil Tseng explained how sales taxes are applied to expenditure reports during the data production process and Geoffrey Paulin described the proper use of imputed income data and the proper use of sample weights in computing population estimates. The latter session noted that the proper use of weights requires a special technique to account for sample design effects that, if not employed, result in estimates of variances and regression parameters that are incorrect.

The afternoon session included two research presentations. First, Brian Melzer, of Northwestern University, investigated the effects of mortgage debt overhang on housing investment and demonstrated that, although homeowners with negative equity cut back substantially on mortgage principal payments, home improvements, and home maintenance spending, these households showed no difference in durable spending on automobiles, furniture, and home appliances—investments that are not attached to the home. In the second presentation, Sayeh Nickpay, of the University of Michigan, described work investigating the seasonality of medical expenditures and flexible spending accounts. Next, Anthony Damico, of the Kaiser Family Foundations, demonstrated a programming code used in the computation of health care expenditures by Medicare households.

The 2011 workshop concluded with CE program staff soliciting feedback from the participants.

2012 workshop

The next microdata users’ workshop will be held July 18–20, 2012, and will be free of charge to all participants, although advance registration is required. For more information about this and previous workshops, visit the CE website (www.bls.gov/cex)
look for “Annual Workshop” under the left navigation bar entitled “PUBLIC USE MICRODATA.”

**Abstracts of presentations**

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

**Jonathan Fisher,** U.S. Census Bureau, “Household Adjustments to Gasoline Price Changes” (Interview and Diary Surveys), day one.

This paper uses aggregate U.S. data, along with household data from the Consumer Expenditure Survey, to explore the adjustments households make to changes in gasoline prices. The paper begins by using aggregate data to show long-run trends in gasoline prices and various measures of gasoline demand. Although gasoline prices have been volatile, the share of total expenditures spent on gasoline has been relatively insensitive to prices, but the level of gasoline expenditures has moved in tandem with gasoline prices.

To investigate how households respond to changes in gasoline prices, the paper then moves to an analysis of household data from the Consumer Expenditure Diary and Interview Surveys. The analysis confirms the well-known finding that the demand for gasoline is inelastic and relatively stable over time and, using the Almost Ideal Demand System of Angus Deaton and John Muellbauer, documents how households alter their nongasoline consumption due to changes in gasoline prices. The paper’s chief finding is that households decrease expenditures primarily on motor vehicles and services.

**Jeff Lundy,** Ph.D. candidate, University of California at San Diego, “Measuring Annual Change in Household Wealth with the CE Survey” (Interview Survey), day one.

The Consumer Expenditure Survey tracks the value of assets and liabilities of a large rotating sample of U.S. households. Unfortunately, researchers studying household wealth have largely neglected this potential resource, relying instead on aggregate statistics. Although aggregate wealth statistics are suggestive of individual household decisions, the Consumer Expenditure Survey has the potential to offer a more direct picture of how U.S. households manage their finances.

To validate the survey’s potential for measuring changes in household wealth, this paper compares the Consumer Expenditure Survey with the well-established Flow of Funds Accounts. Results indicate that the Consumer Expenditure Survey measures change in wealth well at the household level.

In addition, the paper examines the extent of gains and losses in wealth for the 2004–2009 period. Among the findings of the research are that the number of households with annual wealth losses is considerably higher than the number of households with lifetime wealth losses and that wealth gains vary substantially across households in various asset ownership groups. These demonstrative findings reveal the potential of the Consumer Expenditure Survey in examining how households’ financial and demographic characteristics affect their annual change in net wealth.

**Adam Bee,** Ph.D., University of Notre Dame, “The Effect of Car Ownership on Employment: Evidence from State Insurance Rate Regulation” (Interview Survey), day two.

Various economic theories suggest that one reason for low rates of employment among low-skilled, inner-city residents is that they are spatially separated from jobs that have moved out to the suburbs. To test that hypothesis, this paper exploits the variation in state regulation of prior approval of insurance rates. More regulation has been shown to suppress auto insurance prices, thereby decreasing the cost of owning a car. The analysis finds that rate regulation increases the proportion of multicar households, among married couples with children. In those households, the additional car encourages mothers to decrease their labor supply while their husbands increase theirs. One possible explanation of this finding is that second cars are stronger complements to time spent in home production (especially child rearing) than they are to time spent in the labor market.

**Megumi Omori,** assistant professor of sociology, Department of Sociology, Social Work, and Criminal Justice, Bloomsburg University, “What to Buy, When to Buy, and How Much to Spend: Gift Purchasing Between Black Friday and Christmas Eve in 2009” (Diary Survey), day two.

Gift giving has been identified as serving important functions in our society. For example, it shows one’s identity, taste, and status. Gift giving is also a sign of love, and it maintains relationships among family members. Furthermore, it creates boundaries between and within social groups. Gift-giving rituals are common during the holiday season, and it seems that almost everyone is involved in gift shopping. Although several studies have examined gift-giving behavior, such as what to give, who does the shopping, and how much to spend, most of them employed rather small, regionally limited samples.
To my knowledge, there is only one study that used nationally representative data to examine expenditures on gifts: using the 1984–1985 Consumer Expenditure Interview Survey, Thesia Garner and Janet Wagner found that 90 percent of their sample reported some gift expenditures. Among those who spent money on gifts, sample households allocated 3.7 percent of their total expenditures on gifts for other than household members. The authors also found that family size, total expenditures, and education of the reference person correlate positively with gift expenditures.

This paper proposes to update Garner and Wagner’s study by using the most recent 2009 Consumer Expenditure Survey (CE). Differing from their study, it utilizes the CE Diary Survey Expenditure Files and has five objectives:

1. To estimate the proportion of U.S. households that purchase gifts for other than their own household members (hereafter, simply, gifts) and compare that proportion with figures used in other studies (Variable: GIFT).

2. To estimate expenditures on gifts and the proportion of gift expenditures to total expenditures (Variable: COST).

3. To identify the day and date of gift purchasing between Black Friday and Christmas (Variable: QREDATE).

4. To identify items purchased as gifts during the same period (Variable: UCC).

5. To compare the cost of items purchased as a gift and for household members (Variable: COST) and for other than household members, by household characteristics (FMLD files).

Brian Melzer, assistant professor, Department of Finance, Northwestern University, “Effects of Mortgage Debt Overhang on Housing Investment” (Interview Survey), day three.

Homeowners with negative equity have less incentive to invest in their property. They face a debt overhang: some value created by equity investments in the property is expected to go to the lender. Using rich microdata on household expenditures, this paper shows that debt overhang plays an important role in household financial decisions. Specifically, it finds that homeowners with negative equity cut back substantially on mortgage principal payments, home improvements, and home maintenance spending. At the same time, these households show no difference in durable spending on automobiles, furniture, and home appliances—investments that are not attached to the home. The decline in mortgage principal payments is particularly large for negative-equity homeowners in nonrecourse states, where strategic default is more likely because lenders have a claim, albeit limited, on nonhousing wealth. Debt overhang, rather than financial constraints, best explains this set of facts.

Anthony Damico, The Kaiser Family Foundation, “Health Care Spending by Medicare Recipients: How to Analyze CE Data by Expenditure Category for Any Population Group of Interest” (Interview Survey), day three.

This presentation is designed to teach users how to define any population of interest from among the CE Interview files and then, in three steps, rapidly produce line, bar, or pie charts about any expenditure category of interest. First, the presentation walks through how to narrow expenditure categories to only the ones of interest. Second, by making some minor edits to the SAS program included with the dataset (see SAS
program folder, “Intrvw Mean and SE.sas”), one can limit the output to only the expenditures of interest, with the output broken out by any demographic group that one can identify by means of the family files. Third, in order to increase the number of ways to identify demographic groups, the presentation reviews how to merge the family files with some of the other interview files. After completing these 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. Here, the example used is again household health care expenditure categories, but now among Medicare beneficiaries, broken down by all sorts of different demographic groupings. To add one more analytic “trick,” the presentation briefly reviews a technique that can be used to quickly create an “all other” expenditure category, a category containing multiple categories.

BLS speakers

**Staff of the CE program**

William Hawk, Economist, Branch of Information and Analysis (BIA); day one
Craig Kreisler, Economist, BIA; days one and two
Steve Henderson, Supervisory Economist, Chief, BIA; days one (welcoming remarks) and two
Troy Olson, Supervisory Economist, Chief, Phase 3 Section, Branch of Production and Control (P&C); day two
Bill Passero, Senior Economist, BIA; all days
Laura Paszkiewicz, Senior Economist, BIA; all days
Geoffrey Paulin, Senior Economist, BIA; day three
Neil Tseng, Senior Economist, P&C; day three

**Other BLS speakers**

Brian Baker, Technical Writer–Editor, Office of Publications and Special Studies (OPIUBSS), *Monthly Labor Review* Branch; day two
Catherine Hackett, Mathematical Statistician, Division of Price Statistical Methods; day two
Tian Luo, Economist, Office of Field Operations, Western Information Office in San Francisco, Division of Economic Analysis and Information; day one
Kara Markley, Supervisory Economist, Office of Field Operations, Mid-Atlantic BLS Information Office in Philadelphia, Division of Economic Analysis and Information; day one
Karen Ransom, Supervisory Economist, Office of Field Operations, Southeast BLS Information Office in Atlanta, Division of Economic Analysis and Information; day one
Terry Schau, Managing Editor, OPIUBSS, *Monthly Labor Review* Branch; day two

**Speakers from outside BLS:**

Adam Bee, Ph.D., University of Notre Dame, “The Effect of Car Ownership on Employment: Evidence from State Insurance Rate Regulation” (Interview Survey), day two
Anthony Damico, The Kaiser Family Foundation, “Health Care Spending by Medicare Recipients: How to Analyze CE Data by Expenditure Category for Any Population Group of Interest” (Interview Survey), day three
Jonathan Fisher, U.S. Census Bureau, “Household Adjustments to Gasoline Price Changes” (Interview and Diary Surveys), day one; presentation on computation and use of longitudinal weights (Interview Survey), day three
Jeff Lundy, Ph.D. candidate, University of California, San Diego, “Measuring Annual Change in Household Wealth with the CE Survey” (Interview Survey), day one
Brian Melzer, assistant professor, Department of Finance, Northwestern University, “Effects of Mortgage Debt Overhang on Housing Investment” (Interview Survey), day three
Sayeh Nikpay, Ph.D. candidate, University of Michigan, “Seasonality of Medical Expenditures and Flexible Spending Accounts” (Interview Survey), day three
Megumi Omori, assistant professor of sociology, Department of Sociology, Social Work, and Criminal Justice, Bloomsburg University, “What to buy, When to Buy, and How Much to Spend: Gift Purchasing Between Black Friday and Christmas Eve in 2009” (Diary Survey), day two

**NOTES**

1 To preserve the confidentiality of the data, values for some variables, such as income sources and certain expenditures (rent, among others) are topcoded. In this process, values that exceed a predetermined critical value are replaced with a new value. In each case, changed values are flagged for user identification. Details about topcoding are provided in the public-use microdata documentation for the year of interest. (See, for example, 2010 Consumer Expenditure Interview Survey, Public Use Microdata User’s Documentation, September 27, 2011, [http://www.bls.gov/cex/2010/csxintvw.pdf](http://www.bls.gov/cex/2010/csxintvw.pdf).)

2 Coresearchers of the work are Jacqueline Midkiff, of the Mountain–Plains BLS Information Office in Kansas City, and Cheryl Abbot, of the Southwest BLS Information Office in Dallas, both of whom attended the workshop.

3 Luo used CE data in his research into educational expenditures for a project conducted while he was an undergraduate at the University of California at Berkeley.

4 See appendix A for the definition of reference person and consumer unit.

5 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 by 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 about the consumer unit as a whole, such as its 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. The MEMB files in the Interview Survey contain information about each member of the consumer unit, such as the member’s age, ethnicity, and educational attainment. However, the files contain no information on expenditures, because such information pertains to the consumer unit as a whole and therefore is not available on specific members, except in single-member consumer units. The MTAB files in the Interview Survey include information on expenditures at highly 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 detailed levels (e.g., apples, bananas, oranges, which together aggregate to form the summary variable “FRSHFRUT” in the Diary Survey FMLY file). There are also EXPN files in the Interview Survey, which contain even more detailed breakdowns for certain expenditures, and other detailed information on some items, such as the number of members in the consumer unit who are covered by a particular health insurance policy. Some of the differences among all these files were discussed during the overview of the file structure delivered on the first day of training.

In the Interview Survey, the 3-month recall period may include expenditures made in the previous year. For example, persons interviewed in February will report expenditures occurring in November and December of the previous 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 3 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 the previous year’s expenditures and the January expenditures only with the current year’s expenditures.


8 The CE sample design is pseudorandom. However, the proper use of weights requires the use of the method of balanced repeated replication.


APPENDIX A: About the CE data

Consumer unit. The basic unit of analysis in the CE is the consumer unit. In general, a consumer unit consists of any of the following: (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; (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 the three major expenditure categories, either entirely or in part.

Collection and methodology. Since 1980, the Quarterly Interview and Diary Surveys have been collected on an ongoing basis. The Quarterly Interview Survey is designed to collect expenditures for big-ticket items (e.g., major appliances; cars and trucks) and recurring items (e.g., payments for rent, mortgage, or insurance). Some expenditures, such as food at home, are collected globally. In addition to information on expenditures, demographics, and income, information about assets and liabilities is collected. In this Interview 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. Since April 2006, about 7,000 consumer units have participated in the Interview Survey each quarter.

In the Diary Survey, participants record expenditures daily for 2 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 participate annually. Because they complete a separate diary each week, approximately 14,000 diaries are collected each year.

Reference person. The reference person for the consumer unit is the first person mentioned when the respondent is asked to name the person or persons who own or rent the home.

Notes to Appendix A

1 That is, the respondent is asked to provide an estimate of total expenditure for these items, rather than collecting detailed information on items composing the category.

2 A bounding interview collects information aimed at alerting the interviewer to probe in cases where the purchase of a big-ticket or infrequently purchased item reported in one interview is reported, perhaps inadvertently, in the next interview. For example, if the respondent reports purchasing a refrigerator in the first interview and also reports such a purchase in the second interview, the interviewer can ask followup questions to ascertain whether the refrigerator reported in the second interview was the refrigerator 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.
In July 2006, the CE program conducted the first in a series of annual workshops. Held each year in the conference facilities of the Bureau of Labor Statistics (BLS) headquarters in Washington, DC, the workshop has included speakers demonstrating features of the data, as well as reports from researchers who have used these data in their work. Over time, the format has changed to incorporate suggestions from participants, but the basic elements have remained intact.

The first three workshops, held in 2006–2008, took place over 2 days and included concurrent training sessions for novice users and intermediate or expert users. In July 2009, the program was expanded from 2 days to 3 days. The first day was designed especially for new users, including 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 participants as possible. That is, any attendee could participate in 1, 2, or all 3 days of the workshop and benefit from sessions geared toward his or her level of expertise.

Starting with the next workshop, in July 2010, the format was changed slightly. On the basis of comments from the 2009 workshop, research presentations were spread out over the 3 days. Nevertheless, the training and data-descriptive sessions continued to be organized progressively, so that attendees could still plan to attend a combination of days appropriate to their levels of expertise in using the data. In addition, researchers did an excellent job presenting not just results of their work, but processes used, problems or data limitations encountered, how they were handled, and other practical considerations.

Finally, a new feature called “Meet with an Expert” was initiated in the 2010 workshop. 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 use of the data. Several participants did so then and in the 2011 workshop.