

Consumer Expenditure Survey Methods Symposium and Microdata Users' Workshop, 2015

This report describes the fourth annual Consumer Expenditure Survey (CE) Survey Methods Symposium, which took place on July 14, 2015, and the 10th annual CE Microdata Users' Workshop, which took place on July 15–17, 2015.

The CE is the most detailed source of expenditures, demographics, and income collected by the federal government. Every year, the Bureau of Labor Statistics (BLS) CE program releases microdata on the CE website from its two component surveys (the Quarterly Interview Survey and the Diary Survey), which are used by researchers in a variety of fields, including academia, government, market research, and other private industry areas.¹

In July 2006, the CE program office conducted the first in a series of annual workshops to (1) help users better understand the structure of the CE microdata; (2) provide training in the uses of the survey data; and (3) promote awareness of the different ways in which the data are used and explore possibilities for collaboration through presentations by current users and interactive forums. Starting in 2012, the program office added an additional day to the event for a symposium to explore topics in survey methods research in support of the Gemini Project, a major effort to redesign the CE survey (more information here: <https://www.bls.gov/cex/geminiproject.htm>).

In addition to the CE program staff, workshop speakers have included economists from BLS regional offices and researchers not affiliated with the BLS; similarly, symposium speakers have included CE program staff, other BLS National Office staff, and speakers from outside the BLS.



Geoffrey D. Paulin

paulin.geoffrey@bls.gov

Geoffrey D. Paulin is a senior economist in the Consumer Expenditure Survey Program, U.S. Bureau of Labor Statistics.

Nhien To

to.nhien@bls.gov

Nhien To is an economist in the Consumer Expenditure Survey Program, U.S. Bureau of Labor Statistics.

Survey methods symposium

As in previous years, the goals of the 2015 CE Survey Methods Symposium were (1) to provide an update on the status of the Gemini Project, including results from recent projects; and (2) to feature research related to the redesign, including imputation of assets and liabilities, testing the feasibility of collecting outlet data with expenditures, and building a respondent burden index. There were two sessions, one focusing on each topic.

Gemini Project to redesign the CE

Overview of the Gemini Project. As a continuation of the work presented at the 2013 and 2014 CE Survey Methods Symposium, Laura Erhard (CE) provided an overview of the multiyear Gemini Project, which was launched in 2009 to research and develop a redesign for the CE. Much research has been conducted, including recommendations from the National Academies' Committee on National Statistics (CNSTAT). Erhard shared an illustration of the 2013 survey redesign plan and explained the details of the new protocol that were being assessed in the Proof of Concept (POC) Test from July 2015–October 2015. The objectives of the POC Test were to address methodological issues and evaluate operational and experiential factors.

Following the POC Test, CE will be conducting two additional major field tests: an Incentives and Outlets Test in 2016 and a Large-Scale Feasibility Test in 2019. The Incentives and Outlets Test will be conducted in the CE production sample, follow the proposed Gemini Redesign structure of incentives, and look at both operational issues and effects on data quality and response rates. The Large-Scale Feasibility Test will include all the components of the redesign and incorporate lessons learned. It will be conducted much like the POC Test, but with a larger sample, and the instrument will closely resemble the final design.

Results from the Individual Diaries Feasibility Test. Ian Elkin (CE) shared preliminary results from the Individual Diaries Feasibility Test (IDFT), which was designed to inform the operational and data quality aspects of collecting expenditures from household members using personal electronic diaries. The IDFT tested two separate web instruments: a mobile version (via smartphone) and a desktop version. The test was fielded from August 2014–December 2014, and the sample targeted area mobile usage, Internet penetration, multiperson households, and English-speaking households (as a Spanish diary was not developed to reduce costs, although one will be developed for production and included in the Large-Scale Feasibility Test).

The preliminary results reinforced the benefits of electronic diaries from a performance and data quality standpoint. For example, diary keeping information that is automatically generated allows for protocols to be updated more rapidly, electronic diaries allow for the collection of data from paper diaries that are not successfully picked up, and access to up-to-date paradata allows for on-demand interviewer intervention. In addition, an examination of the characteristics of respondents showed that electronic diaries, specifically mobile diaries, do an effective job of targeting groups generally underrepresented in CE data.

Of note, development of CE's electronic diaries is an ongoing, iterative process, along which the IDFT was one milestone. In addition to substantive development milestones (e.g., instrument specifications), protocol milestones also exist, a number of which were omitted from the IDFT fielding. These include burden-appropriate login procedures, as well as specified monitoring and follow-up activities for field staff. Continued testing, into

the Large-Scale Feasibility Test, will bring together the cumulative efforts of substantive development and protocol implementation.

Progress on the CE Electronic Diary. In 2013, CE began the process of building a mobile survey instrument to collect diary data. This work was done to build the instrument used in the IDFT, but extended past the IDFT into developments of later designs. The process began with rough drawings of instrument screenshots that were turned into formal written requirements and handed off to programmers. During the process, a series of usability tests were conducted. Brandon Kopp (Office of Survey Methods Research, OSMR) oversaw and conducted the usability tests and shared an overview of the process along with results and recommendations from the tests at the symposium. Most of the tests focused on the mobile version of the diary, but a subset of tests also looked at the desktop version.

The test results uncovered issues ranging from the login process and password change requirements to respondent difficulty in entering data that met data requirements. Many changes were made to the instruments, but because of limited time and resources, not all issues were addressed. Work to improve the electronic diary is ongoing. There is a plan to propose alternative designs for the desktop version of the diary and conduct usability testing. In addition, CE will be exploring alternate ways of simplifying the login process. Incremental changes will be made up until (and after) it is being used in production.

Gemini Redesign related research

Overview of CE Research. Branch of Research and Program Development (BRPD) Chief Adam Safir began the second session of the symposium with an overview of ongoing CE survey improvements since 2003 and summarized the current research agenda. The agenda, which BRPD updates annually, focuses on research issues within the context of the CE's long-term goals. It also communicates CE's research plans and priorities with respect to the redesign and reflects discussions with internal and external stakeholders. The agenda isn't set in stone and can change over time as new research findings and questions emerge.

Investigating the Imputation of Assets and Liabilities in the CE Interview Survey. Geoffrey Paulin (CE) spoke about the problem of nonresponse and how it affects the data collected on expenditures, income, taxes, and assets and liabilities. While there are methods in place for handling this problem with most of these data, nonresponse to questions that collect data on assets and liabilities is currently under investigation. The purpose of the project is to design a method to impute missing Interview asset and liability data, leveraging models from income imputation and other relevant procedures. The goal of the project is to implement this method into production with 2018 Quarter 2 data.

The team considered several methods, including: one used by the Survey of Consumer Finances, which utilizes an iterative, multiple imputation process; regression trees; and hot deck, but none were feasible. Ultimately, the team decided to investigate a system based on income imputation processing. Paulin described what that system entails and discussed the challenges involved in adapting it to assets and liabilities. The project is a work in progress.

Testing the Feasibility of Integrating Outlets into the CE Diary. Currently, CE and TPOPS (Telephone Point-of-Purchase Survey) collect complementary and potentially redundant information for the CPI (Consumer Price

Index). As a result, the CE and CPI programs are interested in determining the impact on data quality and respondent burden of collecting outlet data in the CE Diary. Erica Yu (OSMR) conducted a small study looking at two ways of integrating outlets in the CE Diary and found that (1) the collection of outlet information did not substantively affect CE data quality, (2) participant ratings of burden showed no large effects due to the addition of outlets, and (3) there was a possible increase in time taken to enter items in the diary. This was a small study, and further research is needed.

Developments in Building a Respondent Burden Index. Danny Yang (OSMR) discussed the work being done to develop a composite burden index for CE that would track perceived respondent burden over time. This would allow CE to detect and understand changes in burden following modifications to the survey, evaluate the association between the survey burden index and other survey measures of interest, and develop interventions that would reduce respondents' perception of burden. Burden scores could be integrated into the overall assessment of survey performance for CE management.

Conclusions: With many research projects underway involving the overall CE Gemini Project, the 2015 CE Survey Methods Symposium was a successful event focused on sharing recently completed and current work with data users and others interested in CE's survey research. These research projects help the program move toward achieving its overall redesign goals, and the symposium serves as a channel for discussion and the exchange of ideas.

The symposium drew a little over 50 attendees from areas such as universities, academic programs in survey methodology, nonprofit organizations, private companies, medical-related establishments, and other federal agencies.

Microdata users' workshop

Day one. The first day of the 2015 workshop opened with presenters from the CE program. Bill Passero provided an overview of the CE, featuring topics such as data collection and publication. Brett Creech then presented an introduction to the microdata, including an explanation of their features, including data file structure and variable naming conventions.

The morning concluded with presentations by researchers not affiliated with the CE program who have used the microdata for a variety of purposes. The first speaker, Stephen Brumbaugh, discussed automobile loans made to low-income consumers. The second speaker, Taylor Smith, related spending patterns to changes in housing wealth in recent years.

After the lunch break, CE economist Aaron Cobet described forthcoming changes in the Public Use Microdata (PUMD) website, and solicited comments from the attendees. The rest of the afternoon was dedicated to practical training, in which attendees had the opportunity to perform programming exercises using the microdata.

The day concluded with an information-sharing group session among workshop participants and CE program staff. This was an open forum in which attendees met informally to discuss their research and offered suggestions for improving the microdata. One recommendation was that the CE make information more readily

available to users. Specifically, the CE needed to find a better way of presenting the documentation, highlighting key topics, and making online help tools more dynamic.

Day two. The second day opened with more advanced topics, with Brian Nix of the BLS Division of Price Statistical Methods presenting technical details about sampling methods and construction of sample weights. Meaghan Smith (CE) followed with a presentation on imputation and allocation of expenditure data in the CE.

The remainder of the morning was dedicated to research presentations by non-BLS attendees. The first of these, entitled “The 2011 Payroll Tax Cut and Household Spending: Evidence from a Quasi-Natural Experiment” (Naveen Singhal), examined how expenditures made by consumers changed in different states in response to a cut in payroll taxes. The experiment became possible because one state, Illinois, raised income taxes by about the same percentage as the reduction in payroll taxes, yielding no net cut for residents of that state. The presenter of the second work, entitled “Household Consumption Smoothing between Monthly Housing Payments” (Li Zhang), had returned to the workshop for a second consecutive year, having presented a different paper (“The Effect of Casinos on Household Consumption”) in 2014. The third presentation, entitled “Income-Expenditure Elasticities of Less Healthy Consumption Goods” (Adam Hoffer), used data from the Diary Survey to analyze expenditures on foods like cola and donuts to estimate how tax increases might affect expenditures on these goods.

After a break for lunch, Carol Boyd Leon and Charlotte Irby, technical writer-editors of the *Monthly Labor Review (MLR)*, described the publication process, from submission to printing, for attendees interested in having their work appear in the *MLR*.

After this description of the *MLR* process, the technical instruction resumed with a presentation of a topic of perennial interest to CE microdata users: how to apply longitudinal weights to the interview data. Following this presentation, Evan Hubener (CE) led a discussion highlighting some of the limitations of the CE survey data and provided best practices for dealing with weights under these circumstances. Hubener detailed how the Interview Survey collects data from respondents for 4 consecutive calendar quarters. During each interview, the respondent is asked to provide information on expenditures for various items during the past 3 months. However, not all participants remain in the sample for all four of these interviews. Those who do remain have different characteristics (e.g., higher rates of homeownership and average age) than those who do not. Therefore, attempting to analyze average annual expenditures by only examining respondents who participate for all four interviews yields biased results.

Following the Hubener presentation, the workshop pivoted to a session explaining an important feature of certain variables in the microdata: topcoding. In a presentation entitled “Balancing Respondent Confidentiality and Data User Needs,” Arcenis Rojas (CE) explained that, in order to preserve the confidentiality of the data, values for some variables, such as income sources and certain expenditures (e.g., 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.² At the conclusion of this presentation, practical training resumed for the rest of the afternoon.

Day three. On the final day, CE staff featured advanced topics, starting with Barbara Johnson-Cox explaining how sales taxes are applied to expenditure reports during the data production process. Next, Geoffrey Paulin

described the correct use of imputed income data and 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.³ Researcher Walter Lake (Pew Charitable Trusts) followed, describing a user-friendly tool he was developing to allow researchers to obtain time-series estimates from microdata both for demographic groups and detailed expenditures not available in online formats through the CE website.⁴ After a break, Aaron Cobet described the new methods in CE for estimating income taxes paid by consumer units, the amounts for which replace those reported by consumers during their interviews, as these data have been found to be extremely unreliable.⁵ The CE uses the National Bureau of Economic Research TAXSIM program to estimate federal and state income taxes. These new estimates were introduced with the publication of the 2013 annual tables. They represent a major improvement to the quality of the CE after-tax income data. The session concluded with a “sneak peek” of developments for CE microdata by Steve Henderson. In 2015, there were many changes made to the Interview Survey. These included the introduction of new health care questions, the dropping of the first interview or “bounding interview,”⁶ and the implementation of a redesigned sample.⁷ Regarding publications, Henderson noted that detailed data tables, which had been available only on request, would be published online, starting with one at the all-consumer-unit level. In addition, a new higher income table and a new table looking at spending by birth year of the reference person, divided into generations, would be released.⁸

After a lunch break, practical training continued, including a presentation of a computer program available with the microdata for use in computing correct standard errors for means and regression results when using (1) unweighted nonimputed data, (2) population-weighted nonimputed data, and (3) multiply imputed income data, both unweighted and population weighted (Paulin). Finally, attendees were debriefed in a feedback session designed to solicit opinion on how to improve future workshops, CE program outreach, and other topics of interest to attendees. Most of the suggestions were related to methods for raising awareness about future workshops. Some users provided additional outlets for us to post information about the 2016 workshop. Users also mentioned the need for additional sample programming codes.

2016 Symposium and workshop

The next Survey Methods Symposium will be held July 12, 2016, once again concomitant with the 11th annual Microdata Users’ Workshop (July 13–15, 2016). While the symposium and workshop will remain free of charge to all participants, advance registration is required. For more information about these and previous events, visit the CE website (www.bls.gov/cex) and look for “Annual Workshop” under the left navigation bar titled “CE PUBLIC-USE MICRODATA.” For direct access to this information, the link is www.bls.gov/cex/csxannualworkshop.htm. Additional details about previous symposia are available at <https://www.bls.gov/cex/geminimaterials.htm>.

Highlights of workshop presentations

Following are highlights of the papers presented during the workshop, listed in the order of presentation. They are based on summaries written by the respective authors.

Stephen Brumbaugh, Ph.D. candidate, UCLA Department of Urban Planning, “Driven to Poverty? An Analysis of Automobile Expenditures in Low-Income Households” (Interview Survey), day one.

Transportation is the second-largest expense category for American households after housing, and the financial burdens of transportation for low-income households—in particular, the costs of buying, operating, and maintaining a vehicle—are a prominent concern among policymakers and antipoverty advocates. Nonetheless, few researchers have directly examined vehicle expenditures in low-income households. In my dissertation, I attempt to fill this research gap by analyzing Consumer Expenditure Survey microdata. My research is guided by three major questions: whether consumer characteristics like race and education explain differences in vehicle expenditures among low-income households; how the nature of vehicle repair expenditures for these households has changed as automotive technology improves; and whether transit expenditures explain differences in automobile expenditures.

Taylor Smith, Ph.D., Georgia Gwinnett College, “How Do Changes in Housing Wealth Affect Consumption Behavior?” (Interview Survey), day one.

Between 1997 and 2006, the price of the typical American house increased 124 percent. This housing boom and its resulting 2008 bust have been cited as major determinants of changes in household consumption over this period. Using more than 12 years of consumer data merged with several macroeconomic time series, we estimate the impacts of housing wealth on 13 specific expenditure categories and the overall budget formation of Americans. We find that housing market fluctuations during this period were indeed a determinant of consumption change, but only in certain sectors, and that the effects were smaller than some news media and previous literature have suggested. Additionally, we show that effect magnitudes vary greatly across young and old homeowners, and across the housing boom and bust periods.

Naveen Singhal, Ph.D. candidate, University of Illinois at Chicago, “The 2011 Payroll Tax Cut and Household Spending: Evidence from a Quasi-Natural Experiment” (Interview Survey), day two.

In 2011, the federal government reduced the payroll tax rate from 6.2 to 4.2 percent, while at the same time Illinois increased its state income tax rate from 3 to 5 percent. Consequently, Illinois workers were largely unaffected by these tax changes, but workers elsewhere experienced an increase in their take-home income. Using this variation in tax liability, I estimate that for every dollar of tax decrease, household spending increased by about 89 cents, especially on recreation, dining, vacations, clothing, and personal care. Additional analysis indicates that the estimates are unlikely to be biased from Illinois-specific shocks and may therefore be interpreted causally.

Li Zhang, Ph.D. candidate, University of Virginia, “Household Consumption Smoothing between Monthly Housing Payments” (Diary Survey), day two.

This paper studies consumption smoothing of households between monthly payments of mortgage or rent. The paper’s focus on regular payments contrasts with most of the literature, which finds excess sensitivity to regular receipt of income. Using the Consumer Expenditure Survey (CE) Diary Survey from 1998 to 2011, I find that spending on nondurable goods is \$3.34, or 9.0 percent higher per day during the two

weeks following the day when a housing payment occurs, compared with the two weeks prior to that day, which is inconsistent with the consumption smoothing predicted by the life cycle/permanent income hypothesis. This finding is robust to the coincident timing of households' regular housing payments and their regular income arrivals, and suggests that findings in the previous literature of excess sensitivity of consumption to regular income arrivals may in part reflect excess sensitivity to the timing of making regular payments. The increase in biweekly average spending following a housing payment day is larger for households in which the household head has lower educational attainment, larger for households with lower income, and has a U-shaped profile in age of household head. My finding is not fully consistent with existing theories that explain departures from consumption smoothing between regular payments, including liquidity constraints and uncertainty about bank account balances.

Adam Hoffer, Ph.D., Assistant Professor, University of Wisconsin-La Crosse, "Income-Expenditure Elasticities of Less Healthy Consumption Goods" (Diary Survey), day two.

There is a long-running policy debate regarding the use of tax policy to modify consumption choices and health outcomes. Specifically, should taxes be imposed on "unhealthful" foods to discourage their consumption and thereby reduce unhealthy outcomes? Objections to this policy include the positing that such goods are price inelastic (i.e., purchases are not sensitive to changes in prices), so the imposition of taxes (essentially equivalent to increasing prices) would be ineffective. This work examines expenditures for cola and donuts, and finds that the expenditures are income inelastic. Therefore, to the extent that taxes reduce income for purchasers of these goods (that is, if the goods cost more, purchasers have less income to allocate to other goods and services), they do little to discourage consumption of these goods.

Walter Lake, Senior Associate, Research Financial Security and Mobility, Pew Charitable Trusts, "Introducing KIWI: A Stata Package to Explore BLS Consumer Expenditure Data" (Interview Survey), day three.

The BLS Consumer Expenditure Survey Public Use Microdata (PUMD) are a very rich, multifaceted set of data with a wealth of information surpassed only by the complexity of the procedures necessary to extract that information. The technical knowledge required to assemble the data prior to analysis creates a barrier for all but the most advanced users of statistical software packages. Lowering the barriers to entry will increase the number of researchers from a variety of fields that can access and utilize the data. To facilitate this, I have created an add-on package for STATA statistical software that streamlines the process for data aggregation and variable creation. Through the use of a graphical user interface (GUI) with drop-down menus and selection buttons, the user can assemble and analyze PUMD with just a few mouse clicks. The GUI allows the user to weight the variables, run crosstabs, and output basic graphs. Two versions of the algorithm that powers the GUI are available to accommodate different levels of statistical programming prowess. The STATA is very functional but still a work in progress and should be ready for public release within the next year.

BLS Speakers

Staff of the CE Program

Cobet, Aaron. Senior Economist, Branch of Information and Analysis (BIA), days one, two, and three

Creech, Brett. Economist, BIA, day one

Curtin, Scott. Supervisory Economist, Chief, Microdata Section, BIA, day one

Henderson, Steve. Supervisory Economist, Chief, BIA, days one and three

Hubener, Evan. Economist, BIA, day two

Johnson-Cox, Barbara. Economist, Branch of Production and Control (P&C), day three

Passero, Bill. Supervisory Economist, Chief, Processing and Analysis Section, BIA, days one and two

Paulin, Geoffrey. Senior Economist, BIA, day three

Rojas, Arcenis. Economist, BIA, days one and two

Smith, Meaghan. Supervisory Economist, Chief, Phase 3 Section, P&C, day two

Other BLS speakers

Boyd Leon, Carol. Technical Writer-Editor, Monthly Labor Review Branch, day two

Irby, Charlotte. Technical Writer-Editor, Monthly Labor Review Branch, day two

Nix, Brian. Mathematical Statistician, Division of Price Statistical Methods, day two

Speakers from outside BLS

Brumbaugh, Stephen, “Driven to Poverty? An Analysis of Automobile Expenditures in Low-Income Households” (Interview Survey), day one

Hoffer, Adam, “Income-Expenditure Elasticities of Less Healthy Consumption Goods” (Diary Survey), day two

Lake, Walter, “Introducing KIWI: A Stata Package to Explore BLS Consumer Expenditure Data” (Interview Survey), day three

Singhal, Naveen, “The 2011 Payroll Tax Cut and Household Spending: Evidence from a Quasi-Natural Experiment” (Interview Survey), day two

Smith, Taylor, “How Do Changes in Housing Wealth Affect Consumption Behavior?” (Interview Survey), day one

Zhang, Li, "Household Consumption Smoothing between Monthly Housing Payments" (Diary Survey), day two

NOTES

1 The Quarterly Interview Survey is designed to collect data on expenditures for big-ticket items (e.g., major appliances, cars, and trucks) and recurring items (e.g., payments for rent, mortgage, or insurance). In the Interview Survey, participants are visited once every 3 months for 4 consecutive quarters.

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 (e.g., white bread, ground beef, butter, or lettuce).

The CE microdata may be downloaded on the CE website (<https://www.bls.gov/cex/pumhome.htm>).

2 Details about topcoding are provided in the public-use microdata documentation for the year of interest. (See, for example, *Consumer Expenditure Interview Survey, Public Use Microdata, 2013 User's Documentation*, September 10, 2014, <https://www.bls.gov/cex/2013/csxintvw.pdf>.)

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

4 Using the link to a BLS-maintained online tool (<https://data.bls.gov/cgi-bin/dsrv?cx>), users can obtain time-series data for published expenditure categories by predetermined demographic series (e.g., age of reference people under 25, 25 to 34, etc.). The new tool will allow users to select data both at detailed levels (e.g., floor coverings) for different groups (e.g., income quintile, age of reference person, or a cross-tabulation of these items) in nominal or real (i.e., inflation-adjusted) dollars. The new tool also allows users to choose whether to display means by calendar year (consistent with CE publications) or collection year (i.e., the year in which the expenditure information was collected, but not necessarily when the expenditures were made). For example, note that with its 3-month recall, Interview Survey respondents who are visited in January are reporting expenditures that took place in the prior year.

5 For details, see Geoffrey D. Paulin and William Hawk, "Improving data quality in Consumer Expenditure Survey with TAXSIM," *Monthly Labor Review*, March 2015, <https://www.bls.gov/opub/mlr/2015/article/pdf/improving-data-quality-in-ce-with-taxsim.pdf>.

6 The purpose of the bounding interview is to ensure that consumers interviewed more than once do not report expenditures in subsequent interviews for which data have already been collected. As an example, if a respondent in the first interview reports purchase of a refrigerator for \$500 and does so once again in the second interview, the interviewer can make sure that the second-interview report is indeed a new refrigerator, different from the one reported 3 months earlier in the bounding survey.

7 The sample redesign occurs decennially, when certain cities or other areas enter the sample and others leave, based on changes in population or other factors.

8 These new tables were introduced in September 2015 and can be found at: <https://www.bls.gov/cex/csxresearchtables.htm>.

RELATED CONTENT

Related Articles

[Consumer Expenditure Survey Microdata Users' Workshop and Survey Methods Symposium, 2014](#), *Monthly Labor Review*, July 2015.

[Consumer Expenditure Survey Microdata Users' Workshop and Survey Methods Symposium, 2013](#), *Monthly Labor Review*, April 2014.

Related Subjects

[Seasonal adjustment](#) | [Consumer expenditures](#)