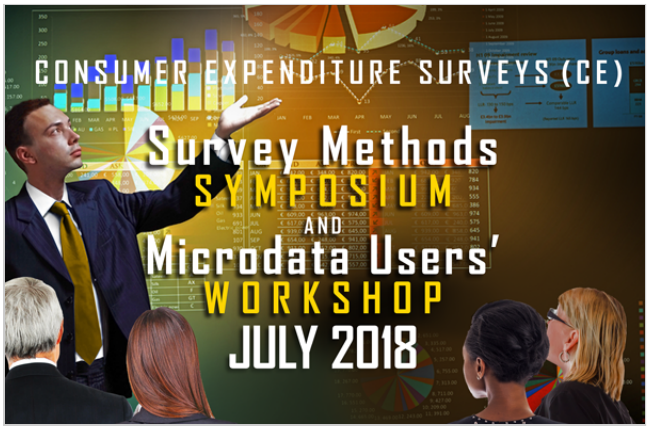
Monthly Labor Review

May 2019

Consumer Expenditure Survey Methods Symposium and Microdata Users' Workshop, July 17–20, 2018

The Consumer Expenditure Surveys (CE) program collects expenditures, demographics, and income data from families and households. The CE program held its annual Survey Methods Symposium and Microdata Users' Workshop from July 17 to 20, 2018, to address CE-related topics in survey methods research, to provide free training in the structure and uses of the CE microdata, and to explore possibilities for collaboration. Several economists from the CE program, staff from other U.S. Bureau of Labor Statistics offices, and research experts in a variety of fields—including academia, government, and private industry—gathered together to explore better ways to collect CE data and to learn how to use the microdata once they are produced.



Geoffrey D. Paulin
paulin.geoffrey@bls.gov

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

Parvati Krishnamurty
krishnamurty.parvati@bls.gov

Parvati Krishnamurty is a senior economist in the Division of Consumer Expenditure Surveys, U.S. Bureau of Labor Statistics.

The Consumer Expenditure Surveys (CE) are the most detailed source of data on expenditures, demographics, and income that the federal government collects directly from families and households (or, more precisely, "consumer units").[1] In addition to publishing standard expenditure tables twice a year, the U.S. Bureau of Labor Statistics (BLS) CE program releases annual microdata on the CE website from its two component surveys (the Quarterly Interview Survey and the Diary Survey). Researchers use these data in a variety of fields, including academia, government, market research, and other private industry areas.[2]

In July 2006, the CE program office conducted the first in a series of annual workshops in order to achieve three goals: (1) to help users better understand the structure of the CE microdata; (2) to provide training in the uses of the surveys; and (3) to promote awareness, through presentations by current users and interactive forums, of the different ways the data are used, and thus provide opportunities to explore collaboration. In 2009, the workshop expanded from 2 days to 3 days to include presentations from data users not affiliated with BLS. This allowed users to showcase their experiences with the public use microdata (PUMD) files (https://www.bls.gov/cex/

pumd.htm), to discuss problems and successes using the data, and to seek comment and guidance from CE program staff in completing their work.

Starting in 2012, the program office has preceded the workshop with a 1-day symposium to explore topics in survey methods research in support of the CE Gemini Redesign Project (Gemini Project), a major initiative to redesign the CE (for more information, go to 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 BLS. Similarly, symposium speakers have included CE program staff, other BLS National Office staff, and speakers from outside BLS. This article describes the 2018 Survey Methods Symposium, conducted July 17, 2018, and the 2018 Microdata Users' Workshop, conducted July 18–20, 2018.

Survey methods symposium

The 2018 Symposium presentations focused on four research topics that are key features of the ongoing Gemini redesign initiative for the Consumer Expenditure Surveys (CE), and followed a similar format to that used in the 2017 Symposium. The four research topics were online diaries, record use, improving data quality through questionnaire design, and innovations in expenditure data collection. The CE program office invited representatives from other federal, international, and private sector surveys to share information about their existing methods and experiences on these research topics. The goals of the symposium were (1) to share CE research findings with stakeholders, survey researchers, and data users and (2) to promote a discussion about common challenges and solutions facing CE and other surveys.

The day was divided into four sessions, with each session centered on one of the four research topics. In each session, a representative from the CE program opened with a presentation on the CE experience, highlighting not only the results of the research, but also the goals to be reached related to the topic and the challenges encountered. The CE presentation was followed by short presentations given by representatives from other surveys on their existing methods or recently completed research relevant to the topic. At the end of each session, the CE representative moderated a discussion about the topic and the presentations, encouraging presenters and attendees to ask questions and provide comments.

In 2018, the symposium drew 65 attendees from universities, nonprofit organizations, private companies, medical-related establishments, and other federal agencies. In the following research topic sections, a review of the presentations is given, followed by a discussion of the key takeaways.

The symposium started with an introduction to the CE redesign by Dr. Parvati Krishnamurty, a senior economist in the CE program at BLS. The presentation outlined the original plans for the redesign, and recent modifications made to the redesign plan for implementation (https://www.bls.gov/cex/gemini-overview-2018.pdf). The redesign plan, which was intended to be implemented in its entirety, was found to have higher costs than the current survey. Therefore, the plan was modified to move to a phased implementation of key design elements into the CE surveys. The phased implementation plan retains the design elements that have been shown to be effective, which include a streamlined questionnaire with less expenditure detail, records focus (including a targeted incentive for record use), online diaries, and token incentives. These elements are to be implemented directly into the CE Diary and Interview surveys. Other design elements—such as a single sample design, two-interview structure, and two-wave

design—have been deferred for testing and implementation in subsequent years, pending changes to requirements or funding availability.

Online diaries

The first session was on online diaries. A major component of the CE redesign plan is the introduction of an online option for respondents to complete the diarykeeping task. This option is an alternative to the existing CE paper-and-pencil diary. The CE is planning a large-scale feasibility test of the online diaries in 2019–20, prior to putting the diary into production.

The good, the bad, and the CE online diary, Ian Elkin (BLS) https://www.bls.gov/cex/good-bad-online-diaries.pdf

Mr. Elkin described CE's journey over the past decade to design an online diary, beginning first with web diaries and then to mobile diaries as an alternative to the paper form that is the only mode currently offered for the CE Diary Survey. During this time, several tests were fielded to assess the usability of the online diary and the feasibility of implementation. These tests included the Web Diary Feasibility Test (2012), Individual Diaries Feasibility Test (2014), Proof of Concept Test (2015), and the Online Diary Improvement Project (2016). Some of these experiments involved comparing the use of personal diaries with household diaries. Since there was evidence that the personal diaries were unpopular with households and interviewers and did not lead to improvements in data quality, personal diaries have been dropped from future tests. Based on findings from past tests, the next step will be to conduct a large-scale feasibility test of online diaries. The online diary—two 1-week household diaries—will be device-optimized so respondents can access the diary on multiple devices including desktops, laptops, tablets, or smartphones. Respondents will be offered a $5 prepaid incentive, but no conditional incentives, for diary keeping. The online diary will have the ability to filter and search entries and sort by expenditure categories. The online diary will also include information on the store, outlet, or website where each item was purchased.

Improving efficiencies on FoodAPS with online food logs, Laurie May (Westat) https://www.bls.gov/cex/foodaps.pdf

USDA's National Household Food Acquisition and Purchase Survey (FoodAPS) is the first nationally representative survey of American households to collect unique and comprehensive data about household food purchases and acquisitions. Detailed information was collected on foods purchased or otherwise acquired for consumption at home and away from home, including foods acquired through food and nutrition assistance programs. FoodAPS collects data using paper diaries on all food spending in a household with the most recent data available to the public, collected from 2012–13. The food categories are similar to those in the CE Diary Survey, but the purpose of the survey differs from the CE Diary, as does its sample design. Ms. May presented results from a pilot test for the FoodAPS that tested the use of an online log of food expenditures that would replace paper diaries. The pilot test involved an online 7-day food diary, available to respondents through an app on their own mobile device or a loaner device provided by Westat. Respondents could access the food diary through a computer with a barcode scanner or via smartphone app. Various features were tested, such as use of a scanner connected to the internet, the use of geocoding to minimize location errors, and respondent uploads of photographs of receipts. The system categorized food-at-home purchases versus food-away-from-home purchases, so that respondents did not have to classify them.

The pilot test was very successful and the technology used was not found to be a barrier to survey cooperation (only 3.5 percent of respondents declined the survey because of concerns with the technology). The data did not require as much editing as those from paper diaries. The average household time spent on the survey was 49 minutes per week. The use of scanning reduced data entry time per item by about 90 seconds compared with manual entry. However, an unexpected finding was that households with numerous purchases used scanning less

than households with only a few purchases. Compared with FoodAPS paper diaries, there was no improvement in item nonresponse and underreporting for the online diary pilot test. The pilot test also showed a drop-off in reporting as the week progressed, which is similar to what is seen for paper diaries. Respondents were able to successfully upload photographs of receipts, but coding these was challenging because receipt structure and naming conventions are not standardized across different types of stores, and very few receipts capture UPCs. Extracting data from receipts in real time has the potential to reduce respondent burden and improve data accuracy. However, receipt scanning software that can capture item-level information is scarce and none can do this task accurately in real time. Moving forward, USDA plans to implement an online diary with more reliance on receipts (as a check on reporting), but continue with manual entry of food purchases and acquisitions into the food log.

Key features of e-diary for the Household Expenditure Survey in Korea, Yeonok Choi (Statistics Korea) https://www.bls.gov/cex/ediary-korea.pdf

Mr. Choi's presentation highlighted the major features of the electronic diary (e-diary) used in South Korea's Household Expenditure Survey. The e-diary was introduced as a way to make the survey less burdensome. Respondents are offered a choice of mode—e-diary or a paper diary. If respondents choose the e-diary, data are automatically transferred monthly; and if they choose a paper diary, data collectors go to the households to collect the monthly paper diaries. Diary keeping is for a 3-month period, with data transmitted to Statistics Korea monthly. An innovative feature of the e-diary is the ability of respondents to link their bank or credit card account data to diaries to facilitate data entry. To alleviate privacy concerns, the survey collector does not have access to general bank data, only to the data that respondents select to be transmitted.

Another feature of the e-diary is a built-in code-search engine to assist diary keepers with item classification, which helps reduce coding error. Additional features include an automated editing system that allows both diary keepers and interviewers to review entries, and online and mobile contact points that facilitate communication between diary keepers and interviewers within the system. These last two features also enable the interviewers to check the presence and progression of e-diary entries made and to prompt diary keepers with reminders throughout the diary keeping period.

For respondents, the diary provides a detailed spending analysis as a nonmonetary incentive. The main advantages of the e-diary are a reduction in both collection cost and nonsampling error. Disadvantages include a higher prevalence of rounding in reporting and incomplete reporting, as well as an adverse effect on survey participation by some older age groups, perhaps due to their unfamiliarity with the technology.

Record use

The second session focused on the use of records to improve data quality. In the CE redesign, this will be done by providing respondents with aids like worksheets and checklists to help them keep track of records; redesigning the streamlined questionnaire to facilitate record use by including introductory wording or separate sets of questions for respondents using records; and providing a monetary incentive for record use.

The CEQ Worksheet: a respondent tool for streamlining the interview experience and improving data quality, Nhien To (BLS) https://www.bls.gov/cex/ceq-worksheet.pdf

This presentation focused on the CE Quarterly Interview Survey (CEQ) worksheet, a respondent tool designed to reduce respondent burden in the CEQ. Input from six census field representatives (FRs) was solicited prior to designing the worksheet in order to learn how they would implement a worksheet based on their own experiences with respondents. The front of the CEQ worksheet includes a brief message to the respondent explaining the purpose of the worksheet; the back of the CEQ worksheet includes several organized tables for respondents to record selected expenses. Based on the feedback solicited from FRs prior to the design, the worksheet includes expenses that are paid by month (such as housing and utilities) and items that are difficult to recall without records

(such as mortgage payments or car loan payments). It also includes a privacy statement requested by FRs. Items on the worksheet are listed in the order they are asked in the CEQ (not in the order of frequency of expenditure) to facilitate the respondent's use of the worksheet during the interview. The worksheet design was revised based on feedback from 52 field representatives. The revised worksheet was field tested to assess the feasibility and effectiveness of the worksheet and its impact on response rates, data quality, and respondent experience. The field test took place in the third quarter of 2018, with worksheets given to 600 respondents after their third interview to help them prepare for their fourth interview. Once data analysis from the test is completed, FRs will be fully debriefed and a report will be prepared. Based on the results of the test, decisions will be made on whether and/or how to implement the Worksheet in the CE Redesign and/or in current production.

The use of respondent records in collecting cost and utilization data on the Medicare Current Beneficiary Survey,* Debra Reed-Gillette *(MCBS) https://www.bls.gov/cex/use-of-respondent-records.pdf

The Medicare Current Beneficiary Survey (MCBS) is a continuous, in-person, multipurpose longitudinal survey covering a representative national sample of the Medicare population residing in the United States and Puerto Rico. Each sampled beneficiary is interviewed up to three times per year for 4 consecutive years to form a continuous profile of their healthcare experiences during their participation in the survey. MCBS collects information on the beneficiaries' health and experiences with the healthcare system and their healthcare expenditures and reimbursements. Respondents' records on their healthcare utilization, costs, and reimbursements are used to provide a total picture of the out-of-pocket costs for their healthcare expenditures.

Ms. Reed-Gillette's presentation described the record-use protocol for the MCBS. Extensive training on understanding the key elements of a large range of healthcare-related records and insurance records is conducted for both interviewers and respondents to aid in record collection. Interviewers, who often have no prior knowledge of the medical field, are trained on how to identify appropriate information from records. In addition, they are taught how to "bundle" a variety of records that relate to a health event that would make up the total cost of an event. During the first interview, the respondent is provided with instructions on the types of records to save for the next interview. Starting with the first interview, the respondent is trained to keep all relevant health-related documents. A planner is provided for respondents to use as a calendar to record all healthcare events, and a folder is provided to use in collecting all relevant records about the events. The information collected from these respondent records is used to estimate the total cost burden and utilization of healthcare for the calendar year for each Medicare beneficiary.

***Does encouraging record use for financial assets improve data accuracy?* Jonathan Eggleston (Census Bureau)** https://www.bls.gov/cex/encouraging-record-use.pdf

Some experimental studies have found no significant reduction in measurement error from the use of records in reporting financial data; however, these studies also had small sample sizes. Dr. Eggleston used financial data from Wave 1 of the 2014 Survey of Income and Program Participation (SIPP) and administrative data from the Internal Revenue Service 1040 Tax Return (IRS 1040) to further investigate the efficacy of record use. The dependent variables in the study were the differences between SIPP and IRS values for each of three types of income—interest income, dividend income, and rental income. The predictors used to proxy for confounding factors in the regression analysis were the respondent's average time spent per question in the SIPP, item nonresponse rate for financial questions (in the SIPP and IRS), and the average amount of rounding in responses to financial questions (in the SIPP and IRS). The record use indicator variable was specific to the asset section of

the SIPP, where the income generated from the assets was recorded. Dr. Eggleston's study found that record use was associated with a 21 percent to 43 percent reduction in measurement error (using IRS 1040s as a benchmark), with only about a 2 percent increase in time to complete the survey (from 41.6 minutes to about 42.5 minutes). Record use was associated with spending an extra 3.5 seconds on each income-from-asset question in the SIPP. About 26 percent of respondents consulted records. Records were more effective at reducing measurement error for rental income as compared with interest income. While the data do not allow for testing various hypotheses, there is some indication that respondents did not usually have earned interest amounts stored in their memory, even during tax season.

Improving data quality through questionnaire design

The third session focused on questionnaire design. Improving data quality is the main goal of the CE redesign. A key element of the CE redesign is the streamlined questionnaire being developed for the CE Interview Survey, which will have more aggregation of items and a record use focus in certain sections.

Revising the CE Surveys to collect outlets, Erica Yu (BLS) https://www.bls.gov/cex/outlets.pdf

A major change being made to the current CE is the addition of new survey questions to collect data on point of purchase (i.e., information about the outlets or businesses where respondents spend money) for selected items in the Interview and Diary Surveys. In this presentation, Dr. Yu summarized the research that was done to prepare for this change and previewed the design of the outlet questions. The Consumer Price Index (CPI) uses a separate random digit dialing telephone point of purchase survey (TPOPS) to identify the stores and businesses where people buy goods and services. However, the TPOPS survey will be discontinued because of high costs and low response rates. The outlet point of purchase questions will be added to the existing CE surveys. The majority of the funds currently spent on TPOPS will be used to increase the CE sample size to provide more assurance of a large enough sample of outlets from CE collection of point of purchase information. Sample size increases for both the Diary Survey and Interview Survey are planned for 2020.

Before proceeding with the inclusion of outlets in the CE, BLS researched the possible impact of adding outlet questions to the CE. BLS expected that adding questions on outlets would not only increase respondent burden, but also might affect the survey experience as a whole. Another concern was that changing the context of the questions to focus on where the item was bought rather than how much was spent would affect data quality. From the CPI perspective: would the new data be comparable to the original source, and would the CE survey yield satisfactory data? Another challenge with adding the outlet questions to the CE (as opposed to collecting the data via TPOPS) is that the TPOPS and CE are different from each other in definitions of item categories, reference periods, sample sizes, and mode, and also have different materials and aids.

Initial exploratory lab studies found that adding outlet questions was feasible and did not negatively affect data quality or perceived respondent burden. Researchers found that the optimal format varies by survey—a transaction-based format worked better for the Diary Survey and an item-based format for the Interview Survey.[3] In 2016, the outlets questions were added on a limited basis to evaluate data quality and objective burden. CE found that the additional outlet questions increased the interview time by about 40 seconds per item category (e.g., televisions, men's suits, or gasoline). However, including them could potentially improve CE data quality by

providing information that could be used during the data review process for remapping expenses that had been put in the wrong category.

Another concern was that CE surveys might yield too little outlet data used by the CPI program for Commodities and Services sample collection design. Exploratory online studies were being conducted at the time of the symposium to test questions and collect more data. Outlet questions are being added to the 2019 Interview Survey that will ask for the name of the business and purchase mode and the city and state where the outlet is located for in-person (as opposed to online) purchases. These questions will be added to a given household's interview on a rotating basis, to limit how many of these questions a respondent gets asked during an interview. For example, one consumer unit would only be asked about outlets for apparel and vehicles, while another would be asked about outlets for entertainment and household appliances. For the 2019 Diary Survey, outlet information is being collected for each selected item reported in an additional column in the diary. For restaurant meals, the Diary Survey emphasizes the "restaurant or vendor" right at the beginning to get a transaction-level report—one entry for the full meal. Outlets are not being collected for clothing, shoes, jewelry, or accessories until the current diary instrument can be redesigned to accommodate this additional field. In the meantime, outlet information for the clothing category will still be collected in the Interview Survey.

MEPS: provider lookup enhancement, Marie Stagnitti (AHRQ) and Angie Kistler (Westat) https://www.bls.gov/cex/meps-lookup-enhancement.pdf

Ms. Stagnitti highlighted the enhanced provider lookup feature used in the Blaise instrument of the Medical Expenditure Panel Survey's (MEPS) Household Component (MEP-HC). There were three goals of the enhanced provider lookup. The first was to reduce the cognitive burden of response, which was achieved by a single string, Google-style search that does not require pre-specified search parameters. The second was to reduce response errors, which was achieved by having an interview-specific, tailored provider directory database reside on the Blaise instrument laptop (CAPI) so that data are pulled directly from the directory. The final goal was to simplify the administration of the interview to lower costs. Ms. Kistler concluded the presentation with a live demonstration of the enhanced provider lookup.

Using historical MEPS data, the MEPS team found 97 percent of medical providers to be within a 100-mile radius of the respondent. Subsequently, they prelimited the scope of the Google-style search by loading only provider names within a 100-mile radius of the respondent's zip code onto the laptops. With the lookup feature, the researchers found a 76 percent match rate for medical providers, which is an improvement from the 60 percent prior to adding the feature. Ms. Stagnitti noted that the match rate could be higher, but interviewers sometimes do not follow through with the search or selection of specific providers, and that this needs to be addressed in training.

Multiphase pretesting during a survey redesign, Mary Davis (U.S. Census Bureau) https://www.bls.gov/cex/multi-phase-pretesting.pdf

This talk focused on multiphase pretesting for survey redesigns, specifically cognitive and usability testing for the National Teacher and Principal Survey. The survey includes questions for principals and teachers about their school and their work. Ms. Davis highlighted how cognitive testing must be an iterative process and does not guarantee a perfect question at the end of the process. As an example, in a question asking teachers about instructional time (which is typically overreported), the exclusionary statement (of what not to include in "instructional time") needed to be pulled into the main question stem. Doing so helped limit overreporting. Ms.

Davis emphasized the importance of iterative cognitive testing of survey questions. She recommended that early rounds of testing need a small number of participants relative to later rounds of testing, as “big problems” tend to surface early. Later rounds need more participants to find more nuanced issues with questions. She recommended doing at least three, but ideally four or five, rounds of cognitive testing.

Innovations in expenditure data collection

In the past few years, there have been many technological advancements in data collection. Some of these innovations, relevant to expenditure data collection, were discussed in the final session of the symposium.

Making audit trails accessible for the CE Quarterly Interview Survey, Brandon Kopp (BLS) https://www.bls.gov/cex/audit-trails.pdf

This presentation summarized work that the CE program has been doing to make audit trail files from the CEQ more accessible to internal researchers. Audit trails are records of all navigation and data transactions within the CE interview survey instrument, every move from one field to another, and every value typed in. These are built into the Blaise programming language, and raw audit trail files are provided for each case for a given month. These audit trails provide information about keystrokes entered into the questionnaire while the interview is being conducted. From these paradata, we can derive information about how the interview was conducted, including time taken to answer each question, changes to the answer, errors, help materials accessed, and whether the interviewer came back to a question later. Audit trails are difficult to work with because, in their native format, they are rows of unstructured text. Mr. Kopp worked on making audit trails accessible by parsing them (i.e., converting these text files into tabular structure—form tables, case tables, field tables, action tables, and error tables). He demonstrated how these tables can be linked and analyzed. The audit trail tables are being developed as a resource for CE staff and are not available to the public.

The use of receipts in the Survey of Household Spending diary, Tom Haymes (Statistics Canada) https://www.bls.gov/cex/use-of-receipts-in-SHS.pdf

Mr. Haymes’ presentation was about the Survey of Household Spending (SHS), the Canadian household expenditure survey. The SHS has both interview and diary components. Fifty percent of the interview sample gets selected for the diary survey. In 2010, the SHS introduced the use of receipts to collect expenditure diary information from households. The option to collect and scan receipts has the potential to reduce burden and increase flexibility for respondents. For the diary survey, respondents can provide transcriptions in a paper diary booklet, provide receipts, or both. Respondents tend to prefer providing receipts when there are a large number of smaller purchases, as with, for example, grocery shopping. Prior to capture, all receipts are manually reviewed by Statistics Canada staff to ensure they fall within the diary reference period, the transaction is approved, and there is no duplication both within receipts and between receipts and diary transcriptions. Diary booklet entries and receipts are scanned into two separate files, and the information is coded at Statistics Canada. The booklet information is captured using optical character recognition, while receipts are currently captured manually from the scanned images. Variation in receipt formats have so far made automatic capture unattainable, but recent technological advances have made automatic capture possible. Illegible receipts are rare and are flagged for imputation. Coding assigns one of more than 650 SHS codes to each item to classify the expenditure. An automated process matches each description to a data dictionary containing common item descriptions with corresponding SHS code. This currently requires an exact match. Items that cannot be automatically coded or

matched are coded manually. Mr. Haymes also highlighted the advantages of using receipts in SHS collection. Respondents who provide both receipts and transcriptions provide the highest amount of expenditures and largest number of items. Another benefit is that less imputation is required for data from receipts compared with transcription.

Do fences make good neighbors? A side by side comparison of RDD and geofencing, Matt Jans (ICF International) https://www.bls.gov/cex/rdd-geofencing.pdf

Dr. Jans presented an example of how nonprobability sampling can be used in conjunction with probability sampling to create population estimates of public health and economic topics. Using the MFour's *Surveys On the Go®* mobile opt-in panel, which includes geofenced[4] grocery, convenience, and liquor stores nationwide, the study aims to obtain representative population estimates using a nonprobability sampling method. It also serves as a proof-of-concept of in-store and in-home image capture as part of a survey protocol. With the Behavioral Risk Factor Surveillance Survey (BRFSS) as a benchmark, this study involves surveying MFour panel members who cross a 50-meter geofence placed around the entrance to the stores used in the study. When a panel member crosses a geofence, regardless of whether they are shopping at that store, the MFour app on their phone makes a "cha-ching" sound like a cash register, announcing a new survey to complete and the incentive amount for the survey. Survey participants are asked to answer about 10 minutes of questions and take an in-store picture of an alcohol, tobacco, or sugar-sweetened beverage display, or other display if those listed are not available. The app also reminds nonrespondents to complete the survey after 1, 24, and 36 hours. Since the MFour panel is slightly skewed toward younger adults and single respondents, ICF International is drawing a census-balanced sample of panel members and making poststratification adjustments to obtain representative estimates. Beyond public health, the study has implications for price and expenditure measurement, for example, by asking respondents to record the prices of items in the store and details about their purchases. More detailed study results are planned for 2019.

Summary of symposium

The CE program office is grateful to the external presenters who shared their experiences on key topics that the CE program is considering. The 2018 Symposium served as a channel for discussing and exchanging ideas to help the CE program move closer to achieving its overall redesign goals. The key takeaways from those discussions for CE include the following:

- An interesting finding from the sessions on online diaries and innovations was the importance of developing an efficient method to capture information from receipts. The FoodAPS online diary app asked respondents to photograph receipts and later coded them, while Statistics Canada has developed a methodology to scan and code receipts. There are technological limitations in automating and coding receipts in real time that need to be addressed in order to improve efficiency. Receipt scanning is something CE is currently investigating, as it has the potential to reduce burden and increase flexibility for respondents, especially if they are making a large number of purchases at a single store.

- Dr. Eggleston's presentation indicated that record use can improve data quality by considerably reducing measurement error for reported asset income collected in the SIPP. However, record use also increases interview length. In addition, record use presents many challenges for both respondents and interviewers. The presentation on MCBS emphasized the importance of training, both for the interviewer and the respondent. Respondents need to know which records to save, how to organize them, and how to draw important information from them. These are important elements for CE to emphasize in the redesign protocols and materials for collecting records-based data in the redesigned streamlined questionnaire.
- The presentations on using apps and geofencing highlighted the importance of GPS and other technological innovations that could enhance the usability and data quality from mobile diaries in the future. For instance, CE could explore the use of apps with GPS or geofencing capabilities to prompt respondents to report expenses in their online diaries when they are at a shopping outlet, or passively collect shopping outlet information for purchases.
- Exclusion wording—instructions to respondents on what *not* to include while answering a survey question—is often used in CE questions. Based on many rounds of cognitive testing, Mary Davis' presentation recommends putting exclusion wording in the main question stem to avoid overreporting. The CE program should keep this in mind as a best practice when developing questions for the streamlined CAPI instrument.

- The Stagnitti-Kistler presentation focused on a database lookup feature in the Blaise instrument that was used in MEPS to look up medical providers. This feature reduced response errors and could potentially reduce cognitive burden for respondents by enabling a single-string, Google-style search of providers. This is relevant for CE because the survey also uses a Blaise instrument. This feature could potentially be used in the CEQ instrument to facilitate collecting the outlet name, insurer name, or other information where the list of options can be customized from a database.

Microdata users' workshop

Meet with an expert: Held in 2017, the 12th annual workshop included an innovation called the "Meet with an expert" program. The purpose was to provide an opportunity for attendees to have in-depth, one-on-one meetings with members of the CE staff, wherein the attendees could ask questions and receive comments or other guidance about the projects in which they were engaged. Attendees were able to sign up for a meeting by checking a box on their registration forms. They could also sign up at the registration desk throughout the workshop. However, the main benefit—both to attendees and CE staff members—of advance registration was to allow the meetings coordinator time to find the most appropriate expert, and time for the expert to investigate the question or prepare other information (handouts, etc.) before the meeting to optimize the quality of the session.

Based on comments from participants, the program was a great success. Therefore, it was repeated in the 2018 workshop. Several attendees arranged meetings by registration form, email, or onsite forms, and the planning team received positive comments on the program. In fact, in a feedback form submitted after the workshop, one participant described it as "the most useful part" of the workshop.[5]

The program is being continued for the 2019 Microdata Users' Workshop. Once again, attendees are able (and encouraged) to arrange meetings via registration form, email, or onsite.

Day one: The first session of the 2018 workshop consisted of presenters from the CE program. After welcoming remarks by Branch of Information and Analysis (BIA) Chief Steve Henderson, Program Manager Adam Safir provided an overview of the CE, featuring topics including how the data are collected and published. Economist Jimmy Choi then presented an introduction to the microdata, including how they can be used in research and the types of documentation about them available to users. Economist Arcenis Rojas completed the session with a description of data file structure and variable naming conventions.

After a break, attendees received their first practical training with the data. In this session, they learned basic data manipulation, including how to compute means from the microdata for consumer units with different characteristics (e.g., by number of children present).

Following a lunch break, Senior Economist Aaron Cobet (BIA) explained the need to balance confidentiality concerns of respondents with usefulness of the data to researchers. Because of Title 13, the U.S. Code that requires confidentiality of response, information that might potentially identify specific respondents must be removed from the CE data before they are released publicly. Some identifiers are direct, such as names and addresses. Others are not direct, such as extremely high expenditures or make and model of automobile(s) owned.

Mr. Cobet explained the methods used in the production of the CE microdata files to address these concerns. The first method, called “topcoding,” involves reported values for income or expenditures that exceed a certain threshold, called the “critical value.” These values are replaced by an average of all values exceeding this threshold and then “flagged” as topcoded (or “bottom-coded,” in the case of large income losses).[6] He also explained recoding, in which data are either made less precise (e.g., if the owned automobile was produced in 1999, the year is replaced with the decade of manufacture [1990s in this example]) or changed in another way (e.g., state of residence is changed from Delaware to New Jersey) to preserve both comparability and confidentiality. Mr. Cobet next explained suppression, in which reported values are removed from the data set. In some cases, only specific information is suppressed on a record (e.g., details of a specialized mortgage). In other cases, the entire record is removed (e.g., report of a purchase of an airplane).[7] Finally, Mr. Cobet talked about methods to eliminate “reverse engineering,” a process through which the user could deduce protected information from other information provided in the publicly available files.[8]

Following this presentation, practical training resumed with a project designed to obtain sample means based on detailed data on educational expenditures derived from various files.[9] Attendees also learned how to integrate results from the Interview and Diary Surveys to match expenditure categories in CE published tables.

Presentations from researchers not affiliated with the CE program completed the afternoon activities. (Note that summaries of the papers presented by outside researchers are included at the end of this Conference Report.)

The first speaker, Ph.D. candidate Rosa Lee (The George Washington University)—the first in the discipline of public policy and public administration to address a CE workshop—spoke about her use of CE microdata to study expenditure patterns of the middle class.

The second speaker in this session was Ting Lan, a Ph.D. candidate in economics at the University of Michigan. Ms. Lan used data from both the Interview and Diary Surveys to assess relationships between monetary policy and consumer spending at different points of the income distribution when prices are sticky.

The final speaker in this session was Dr. Zheli He, an economist with the Penn Wharton Budget Model, an organization housed at the Wharton School of the University of Pennsylvania. Dr. He’s work used CE data to investigate the relationship between the marginal propensity to consume (i.e., the share of each additional dollar that a consumer allocates to a particular good, service, or aggregated set of goods and services) and permanent income (i.e., a function of income actually received today plus expectations of income to be received in the future). Dr. He also included net worth and consumer unit characteristics (age of reference person, etc.) in her estimation of permanent income.

Each of these studies used income in increasing degrees of complexity. For example, the first two presentations used total income before taxes to categorize consumer units into groups (e.g., the “middle class”). The third used a historical series of income after taxes as a dependent variable in regression analysis, the results of which were used to estimate permanent income.

As these presentations demonstrate, the use of income data from the CE is frequent, as the relationship between expenditures and income is self-evident. However, as in many household surveys, income data are subject to nonresponse, which can cause bias in estimates of all types—from simply assigning a consumer unit to the wrong

income category to incorrect parameter estimates in regression. The same caveat applies to tax data, from which estimates of income after taxes are derived.[10]

For this reason, Dr. Geoffrey Paulin, a senior economist in the CE program (BIA) and leader of the CE income imputation team since its inception, served as discussant for this session, marking the first time a workshop session featured a formal discussant. He explained the history of the treatment of income data, noting how publication procedures changed with processing of data collected in 2004, when multiple imputation of income was introduced.[11] He noted that multiply imputed data require special techniques for proper analysis, whether computing a mean, variance, or regression coefficient for a multiply imputed variable such as income in CE. (See "Users' Guide to Income Imputation in the CE," https://www.bls.gov/cex/csxguide.pdf, for details.) He also explained recent improvements in the processing of income tax data. No longer relying on respondents to report values, these data are now (since publication of results collected in 2013) estimated based on reported (or imputed) income before taxes.[12]

Following Dr. Paulin's discussion, the afternoon concluded with a networking opportunity for attendees. The event was an informal gathering both to allow them to meet each other and to initiate or renew contacts with staff of the CE program.[13]

Day two: The second day opened with more advanced topics. First, statistician Brian Nix of the BLS Division of Price Statistical Methods (DPSM) presented technical details about sampling methods and construction of sample weights. Next, statistician Barry Steinberg (DPSM) described a project on which he has worked with co-author Sally Reyes Morales (DPSM) to implement changes to CE consumer unit population weight processing, based on results of the 2010 Census. The concluding presentation of this section featured economist Taylor Wilson (BIA) presenting the introduction of experimental weights for estimating state-level expenditures with the use of the CE microdata. He noted that weights for New Jersey, California, and Florida were available (https://www.bls.gov/cex/csxresearchtables.htm#stateweights). Mr. Wilson also presented the criteria used by the CE division to assess the feasibility of devising weights for other states.

Following a break, practical training resumed. In this training, attendees learned how to obtain information on nonexpenditure characteristics, such as type of school attended, associated with certain educational expenditures, using detailed PUMD files.[14]

They also received an introduction to the procedures needed to obtain consumer-unit-population weighted averages for expenditures; that is, instead of computing mean expenditures from the sample itself, how to apply weights to estimate mean expenditures for the consumer unit population as a whole. [15]

Following this training, Dr. Paulin introduced an informal panel consisting of two economics students, each of whom had used the Diary Survey to study expenditures for food at home.[16] The first speaker was Lacey Wilson, an undergraduate student from the University of South Carolina. Her research examined whether adults of different ages purchase different types of foods based on the USDA healthy eating recommendations that were in effect when they were in their formative years. The second speaker, graduate student Yiting Lan (The Ohio State University), presented her work investigating whether temporary increases in benefits in the Supplemental Nutritional Assistance Program (SNAP) enacted in 2009 were associated with increased purchases of fresh fruits and vegetables.

After a break, Terry Schau, managing editor of the *Monthly Labor Review* (*MLR*), described the *MLR* publication process, from submission to posting, for authors interested in having their work appear in the *MLR*.

Following this presentation, Dr. Paulin described the correct use of sample weights in computing consumer unit population estimates. He noted that the proper use of weights requires a special technique to account for sample design effects that, if not employed, results in estimates of variances and regression parameters that are incorrect. He also mentioned a topic of perennial interest to CE microdata users: caveats concerning the use of data only from respondents who complete all four interviews of the Interview Survey.[17] This led into a practical training session devoted to computing weighted results in two projects: one related to computing results for collection year estimates, and the other for calendar year estimates. The distinction is that collection year refers to the date on which the respondent reported the expenditures to the interviewer while calendar year refers to the period in which they actually occurred. For example, for a person participating in the Interview Survey in January 2017 who reports expenditures that occurred during the final three months of 2016 (October, November, or December), the expenditure collection year is 2017, while the expenditure calendar year is 2016.

The afternoon concluded with two presentations by non-BLS attendees. The first presentation, by Ph.D. candidate Karim Nchare (Penn State University), described empirical work testing the implications of the “normality” assumption in demand for goods. That is, economists define “normal” goods as those for which quantity demanded increases as income increases (given no change in price). An example of why understanding normality is important is to anticipate demand effects of rising or falling prices for a good or service due to policy changes, such as changes in taxes on food. The second presentation, by Dr. Michael Conte (RegionalOneSource), described the development of a website that will combine data from CE and the American Community Survey (a product of the U.S. Census Bureau). The website will enable users to ascertain the buying power (as measured by income after taxes) and the dollar value of spending by consumers on various types of goods and services within user-defined geographical boundaries as large as the U.S. or as small as a census tract or block group.

Day three: The final day started with CE staff discussing advanced topics. First, Economist Barbara Johnson-Cox of the Branch of Production and Control (P&C) explained how sales taxes are applied to expenditure reports during the data production process. Then, Economist Clayton Knappenberger (P&C) spoke about imputation and allocation of expenditure data in the CE. Finally, Taylor Wilson described the efficacy of normalizing expenditure and income data when performing regression analysis to achieve better results.[18] Specifically, the presentation described the use of “power” transformations (e.g., regressing the square root of expenditures on the cube-root of income) to achieve this goal.[19]

Next, a panel of two outside researchers, moderated by Dr. Paulin, addressed research related to transportation expenditures. The first panelist, Dr. David Poyer (Morehouse College/Argonne National Laboratory), previous attendee and first-time presenter, described his work investigating changes in ride-hailing and ride-sharing expenditures, such as taxi or similar services. He was followed by Dr. Jonathan Peters (College of Staten Island/ University Transportation Research Center), who examined how expenditures on transportation are changing given changes in technology (e.g., smartphones with built-in GPS tracking services) and transportation services (new companies and products, such as Uber and Lyft, that compete with traditional taxi services, and Zipcar, that provides extremely short-term rental cars).

The panel was followed by the final presentations of the workshop. The first of these was a “sneak peek” of developments for CE publications and microdata. BIA Chief Steve Henderson noted several developments, including the addition of a new question, scheduled to start in 2019, to ask whether anyone in the consumer unit has previously served in the U.S. military. This question, which will supplement a current question asking whether anyone in the consumer unit is currently serving in the U.S. military, is being added in response to requests from different federal agencies regarding the economic status of U.S. veterans. Also in 2019, the CE will publish data at more refined geographic levels (census division in addition to current census region), and a new column on the (still new) generational tables (first published officially to reflect 2016 data) showing expenditures for the “post-Millennial” generation.[20]

Continuing the “sneak peek” theme, Dr. Paulin described work in progress within the CE program to impute data for assets owned and liabilities owed when the holding, but not specific value, of either is reported.

Following a lunch break, the workshop reconvened for a feedback session led by Dr. Erica Yu (BLS). In the session, attendees had the opportunity to provide comments on what they found most (or least) useful about the workshop, and to make suggestions for future events. The comments were overwhelmingly positive, with attendees agreeing that the balance of training and research presentations, along with the content of these presentations, were appropriate.[21]

The final training session was devoted to the computation of means, standard errors, and regression parameter estimates when using multiply imputed data, such as income in CE. In addition, those interested received an instruction manual for use of a computer program for SAS software users that is available with the microdata. This program helps CE microdata users to compute correct standard errors for means and regression results easily when using (1) unweighted nonimputed data, (2) population-weighted nonimputed data, and (3) multiply imputed income data, both unweighted and population weighted. Finally, a few attendees took one last opportunity to meet with an expert at this year’s workshop.

SYMPOSIUM AND WORKSHOP OF 2019

The next Survey Methods Symposium will be held July 16, 2019, in conjunction with the 14th annual Microdata Users’ Workshop (July 17–19). Although the symposium and workshop remain free of charge to all participants, advance registration is required (https://data.bls.gov/forms/cex-registration.htm). For more information about these and previous events, visit the CE website (https://www.bls.gov/cex/) and under the

left navigation bar, titled “CE PUBLIC USE MICRODATA,” look for “ANNUAL WORKSHOP.” For direct access to this information, the link is https://www.bls.gov/cex/csxannualworkshop.htm. Links to the agendas for the 2018 workshop (https://www.bls.gov/cex/ce-2018-workshop-agenda.pdf) and the 2018 symposium (https://www.bls.gov/cex/ce-2018-symposium-agenda.pdf) are also available on this web page. Both agendas include links to presentations delivered at the respective events.

Highlights of workshop presentations

The 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.

Hyun Kyong “Rosa” Lee, Ph.D. candidate, The George Washington University, “Consumption patterns of the American middle class in major U.S. metropolitan areas” (Interview Survey), day one.

> Does the middle class or upper class have distinctive consumption patterns? Consumption capacity is an important measure of class distinction, just as income data are. Although consumption capacity has been used for a definition of middle class in the developing world, its use for class studies is rare in developed countries, including the United States. Even among scholars who prefer income-based definitions, it is commonly accepted that income measure is not sufficient to operationalize the middle class, since the definition does not consider the consumption of the economic actor, according to various sources in the literature. This project aims to identify distinctive consumption by American households that fall into the middle class—defined as having income between 75 percent and 250 percent of the national median household income. This study utilizes Consumer Expenditure (CE) Survey Public-Use Microdata (PUMD) published by the Bureau of Labor Statistics to see whether middle-class households show distinctive consumption patterns relative to other class categories. The CE Interview Survey is an underutilized data source, which collects 95 percent of the total expenditures and income by households, on the topic of American middle class. Examining detailed data on educational expenses, mortgages and rent expenses, medical and health expenditures, and trips and vacation expenditures in the 2015 survey data, this work contributes to the understanding of how American middle-class households face different burden across the U.S.—focusing on major Metropolitan Statistical Areas. This paper calls for broadening class definitions to include both income and consumption measures to a better standard of living and its impact on regional economic growth.
>
> To address this, I replicate the CE “Average expenditure, share, and standard error” table for the three income categories: below-middle class, middle class, and above-middle class, in the same format in which BLS publishes. I mainly use eight Consumer Price Index Consumption basket categories, plus two more expenditure categories: vacation and personal insurance/pension/savings categories.

Ting Lan, Ph.D. Candidate, University of Michigan, “Price stickiness along the income distribution and the effects of monetary policy” (Interview and Diary Surveys), day one.

This project proposes and quantifies a novel mechanism through which monetary policy shocks have distributional consequences. By using the data from the Consumer Expenditure Survey (CE), we obtain expenditure shares across detailed product categories for households at different percentiles of the income distribution. Combining them with the item-level consumer price data from BLS and the price stickiness constructed by published authors, we document that the prices of goods consumed by high-income households are stickier and less volatile than those of goods consumed by middle-income households. This suggests that monetary shocks can have distributional consequences by affecting the relative prices of goods consumed at different points on the income distribution. We use a Factor-Augmented VAR (FAVAR) model to show that, following a monetary policy shock, the estimated impulse responses of high-income households' consumer price indices are 22 percent lower than those of the middle-income households. We then evaluate the macroeconomic implications of our empirical findings in a quantitative New-Keynesian model featuring households that are heterogeneous in their income and consumption patterns, and sectors that are heterogeneous in their frequency of price changes. We find that: (1) the distributional consequences of monetary policy shocks are large and similar to those in the FAVAR model; and (2) greater income inequality increases the effectiveness of monetary policy, although this effect is modest for realistic changes in inequality.

Zheli He, Ph.D., Economist, Penn Wharton Budget Model, "Marginal propensity to consume out of permanent income" (Interview Survey), day one.

This paper provides an alternative method to calculate the marginal propensity to consume (MPC) for a given gender, race, education, and age group, as well as its variation across individuals within one of these groups. First, we provide a regression framework for analyzing the effects of individual attributes on total family income after tax using the Consumer Expenditure Survey. Second, we use the Panel Study of Income Dynamics to construct transition paths of these attributes over the life cycle, conditional on gender, race, initial education, and age. Finally, we use the estimated regression coefficients and the projected demographic profiles to measure family permanent income. The MPC is calculated by regressing consumption expenditures on income shocks and the permanent component of income, controlling for all attributes. We find that households typically spend 20 cents out of each dollar of income shocks. On the other hand, if one's permanent income goes up by $1, then their consumption expenditures go up by 1.8 cents. Interestingly, we also find that the MPC out of income shocks is not statistically different for people at different permanent income levels. To our knowledge, this paper is the first to take into serious consideration the changing individual characteristics when calculating permanent income, which provides us with a more accurate measure. Furthermore, our results have important implications for polices that aim at increasing aggregate demand based on the assumption that low-income households have a higher MPC.

Lacey Wilson, Undergraduate, University of South Carolina, "U.S. dietary recommendations and grocery spending: a cohort analysis" (Diary Survey), day two.

During the late 1970s and early 1980s, foods with high levels of fat and cholesterol were purported to directly cause heart disease and targeted as unhealthful. Dietary recommendations from both the USDA and popular media during this time emphasized avoidance of these products. Previous studies show that individuals tend to retain beliefs learned during childhood; this study will address whether Americans who

grew up during this period retained a tendency to avoid high fat and cholesterol foods. We use expenditure data from the Bureau of Labor Statistics' Consumer Expenditure Survey to analyze the spending patterns of different generations in the context of dietary recommendations learned during primary school. Respondents are split into groups based on birth year, and we find group averages for percentage of total grocery dollars spent on various food items between the years 1996 and 2014. We find that individuals who were of primary school age during the late 1970s and early 1980s allotted a significantly lower percentage of grocery expenditures to eggs than did their older counterparts. These results imply that Americans who were recommended against consuming high-cholesterol foods during childhood may continue to consume less of those foods than do those who were not, regardless of a later change in those recommendations.

Yiting Lan, Ph.D. Candidate, The Ohio State University, "The impact of in-kind food benefit increase on consumption: evidence from the SNAP" (Diary Survey), day two.

The purpose of this study is to investigate the impact of Supplemental Nutritional Assistance Program (SNAP) benefit increases on participants' purchase of fresh vegetables and fruits. In order to investigate the question, Consumer Expenditure Survey data of 2007 through 2011, when SNAP benefits experienced several large discrete increases, are used. The dependent variables include food at home (FOODHOME), fresh vegetables (FRSHVEG), and fresh fruits (FRSHFRUT). A dummy variable is used to indicate receipt of SNAP benefits in the past month. Demographic variables, including gender, age, race, family size, etc., are used as fixed effects. With the Consumer Expenditure Diary Survey data, the hypotheses include the following:

- Hypothesis1. SNAP households will increase their expenditure for food at home when their SNAP benefits increase. This is represented by an increase of weekly expenditure for food and nonalcoholic beverages purchased at grocery stores.
- Hypothesis2. SNAP households will increase the amounts of healthy foods purchased in the home when their SNAP benefits increase. This is represented by an increase of weekly expenditures for fresh vegetables at home.
- Hypothesis3. SNAP households will increase the amounts of healthy foods purchased in the home when their SNAP benefits increase. This is represented by an increase of weekly expenditures for fresh fruits at home.

To exclude the effect of macroeconomic change and test the impact from a SNAP benefits increase, a difference-in-difference design is used. The difference-in-difference design assumes that the treatment group and nontreatment groups experience similar trends if the treatment does not occur.

Karim Nchare, Ph.D. Candidate, Penn State University, "Testable implications of normality in stochastic demand for two goods" (Diary Survey), day two.

A good is normal if its consumption increases with income, keeping prices fixed. I derive the testable implications of normal demand in a two-goods setting where data are from a repeated cross-section, unobserved heterogeneity is completely unrestricted, and endogeneity of total expenditures is allowed. Using revealed preference restrictions, simple closed-form expressions characterize whether (population

level) data are consistent with the normality assumption. I illustrate the empirical relevance of our theoretical results through an application to data drawn from the CE microdata.

Michael Conte, Ph.D., RegionalOneSource (ROS), "Using the Consumer Expenditure Surveys Microdata to estimate consumer spending and buying power at any level of regionality" (Interview and Diary Surveys), day two.

This presentation describes a project to combine data from two of the country's most powerful survey databases—the American Community Survey (ACS) and the CE Public Use Microdata (CEPUMD)—to provide estimates of consumer spending and buying power for any U.S. geography, ranging from geographies as large as the entire country to those as small as a census tract or census block group. The authors of the research know of no data source that provides estimates of consumer spending in all areas of spending (for example, not just grocery store purchases) at sufficiently small levels of geographical specificity so as to be useful in preparing a typical business plan or a governmental or not-for-profit project plan. The question that this research seeks to answer is whether it is possible to "marry" the data from the ACS with the CEPUMD in order to provide such estimates.

David Poyer, Ph.D., Professor, Morehouse College/Argonne National Laboratory, "Tracking changes in ride-hailing/ride-sharing expenditures" (Interview Survey), day three.

The purpose of this research is to undertake an econometric analysis/assessment of household spending data with the specific objective of determining changes in the composition of transportation expenditures over time. Further, we analyze how changes in ride-hailing/ride-sharing expenditures (as measured by "local taxi and limousine expenditures" UCC code 530412) are dynamically related with other transportation expenditure categories (particularly vehicle expenditures).

Jonathan Peters, Ph.D., The College of Staten Island/ Research Fellow, University Transportation Research Center, "Just what do we actually know about household spending on transportation services and how are they changing in the 21st century" (Interview and Diary Surveys), day three.

U.S. households spend roughly 17 percent of their household income on transportation services. Yet, as BLS Senior Economist Geoffrey Paulin discussed in a recent *Monthly Labor Review* article "Fun facts about Millennials: comparing expenditure patterns from the latest through the Greatest generation" (March 2018), households of various generations are exhibiting different patterns of consumption with respect to transportation services. Millennial households are spending a greater percentage of household income on transportation (18.9 percent of household income) as compared with any other generation. In addition, Millennials own far fewer automobiles (1.5 per household) as compared with Generation X and Baby Boomer households (2.1 and 2.2, respectively). This study looks to expand on prior work by Peters, King, Gordon, and Santiago to explore the component parts of transportation spending in the CE. In particular, we look to further study spending on road tolls, taxi type services, fuel use, mass transit fares, and the overall cost of automobile ownership. Results are segregated by income class, educational status, race, age, and household geographic location. A goal of the research was to understand the component contributors to household consumption patterns.

WORKSHOP PRESENTERS

Staff of the CE program

Choi, Jimmy. Economist, Branch of Information and Analysis (BIA); day one

Cobet, Aaron. Senior Economist, BIA; day one

Curtin, Scott. Supervisory Economist, Chief, Microdata Section, BIA; emcee and practical training sessions; days one, two, and three

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

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

Knappenberger, Clayton. Economist, P&C; day three

Paulin, Geoffrey. Senior Economist, BIA; days one, two, and three

Rojas, Arcenis. Economist, BIA; day one

Safir, Adam. Chief, Division of Consumer Expenditure Surveys; day one

Wilson, Taylor. Economist, BIA; days two and three

Other BLS speakers

Nix, Brian. Mathematical Statistician, Division of Price Statistical Methods (DPSM); day two

Schau, Terry. Managing Editor, Monthly Labor Review; day two

Steinberg, Barry. Mathematical Statistician, DPSM; day two

Yu, Erica. Research Psychologist, Office of Survey Methods Research; day three

Non-BLS speakers

Conte, Dr. Michael (Ph.D.). RegionalOneSource (ROS), “Using the Consumer Expenditure Surveys Microdata to Estimate Consumer Spending and Buying Power at any Level of Regionality” (Interview and Diary Surveys); day two. Prior attendee (2012, 2014, 2016, and 2017) and presenter (2017); returning presenter (2018).

He, Dr. Zheli (Ph.D.). Economist, Penn Wharton Budget Model, “Marginal Propensity to Consume Out of Permanent Income” (Interview Survey); day one. First-time attendee and presenter (2018).

Lan, Ting. Ph.D. Candidate, University of Michigan, “Price Stickiness along the Income Distribution and the Effects of Monetary Policy” (Interview Survey); day one. First-time attendee and presenter (2018).

Lan, Yiting. Graduate Teaching Associate, The Ohio State University, “The Impact of In-Kind Food Benefit Increase on Consumption: Evidence from the SNAP” (Diary Survey); day two. First-time attendee and presenter (2018).

Lee, Hyun Kyong “Rosa.” Ph.D. Candidate, The George Washington University, “Consumption Patterns of the American Middle Class in Major U.S. Metropolitan Areas” (Interview Survey); day one. Prior attendee (2017); first-time presenter (2018).

Nchare, Karim. Ph.D. Candidate, Penn State University, “Testable Implications of Normality in Stochastic Demand for Two Goods” (Diary Survey); day two. First-time attendee and presenter (2018).

Peters, Dr. Jonathan (Ph.D.). The College of Staten Island/Research Fellow, University Transportation Research Center, “Just What Do We Actually Know about Household Spending on Transportation Services and How Are They Changing in the 21st Century” (Interview and Diary Surveys); day three. Prior presenter (2014 and 2017); returning presenter (2018).

Poyer, Dr. David (Ph.D.). Professor, Morehouse College/Argonne National Laboratory, “Tracking Changes in Ride-hailing/Ride-sharing Expenditures” (Interview Survey); day three. Prior attendee (2011, 2015, and 2017); first-time presenter (2018).

Wilson, Lacey. Undergraduate, University of South Carolina, “US Dietary Recommendations and Grocery Spending: A Cohort Analysis” (Diary Survey); day two. Prior attendee (2016); first-time presenter (2018). Note: Ms. Wilson was the first undergraduate student to present at a CE workshop.

SUGGESTED CITATION

Geoffrey D. Paulin and Parvati Krishnamurty, "Consumer Expenditure Survey Methods Symposium and Microdata Users’ Workshop, July 17–20, 2018," *Monthly Labor Review,* U.S. Bureau of Labor Statistics, May 2019, https://doi.org/10.21916/mlr.2019.11.

NOTES

1 Although a household refers to all people who live together in the same living quarters, “consumer unit” refers to the people living therein who are a family, or others who share in specific financial arrangements. For example, two roommates living in an apartment constitute one household. However, if they are financially independent, they each constitute separate consumer units within the household. Similarly, although families are related by blood, marriage, or legal arrangement, unmarried partners who live together and pool income to make joint expenditure decisions constitute one consumer unit within the household. For a complete definition, see the CE glossary at https://www.bls.gov/cex/csxgloss.htm. For more information on households and families, see https://www.census.gov/programs-surveys/cps/technical-documentation/subject-definitions.html#household.

2 The Quarterly Interview Survey is designed to collect data on expenditures for big-ticket items (e.g., major appliances or automobiles) and recurring items (e.g., payments for rent, mortgage, or insurance). In the Interview Survey, participants are visited once every 3 months for four consecutive quarters. In the Diary Survey, on the other hand, participants record expenditures daily for 2 consecutive weeks. This 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 for both surveys may be downloaded from the CE website at https://www.bls.gov/cex/pumd_data.htm.

3 In a transaction-based format, the respondent reports information for expenditures for all items that might appear on one receipt. For example, apples, bananas, and oranges may have been purchased in one transaction at the grocery store. An item-based format relies on specific things bought, regardless of date or number of transactions; for example, all apples purchased in the last three months, regardless of purchase venue, followed by oranges.

4 Geofencing is the practice of using global positioning (GPS) or radio frequency identification (RFID) to define a geographic boundary. Then, once this "virtual barrier" is established, the administrator can set up triggers that send a text message, email alert, or app notification when a mobile device enters (or exits) the specified area.

5 Since 2017, attendees who were either unable to attend the feedback session in person or who have had comments to share later were encouraged to provide them electronically. See "Request for Comments," https://www.bls.gov/cex/ceworkshopthankyou.htm.

6 For example, suppose the threshold for a particular income or expenditure is $100. On two records, the reported values exceed this: $200 on record A and $600 on record B. In this case, the value is topcoded to $400 (the average of $200 and $600), and the reported amounts are replaced with $400. An additional variable, called a "flag," is coded to notify the data user that the $400 values are the result of topcoding, not actual reported values.

7 For details on topcoding and suppression, including specific variables affected and their critical values, see "2016 Topcoding and Suppression," August 29, 2017, https://www.bls.gov/cex/pumd/2016/topcoding_and_suppression.pdf. Additional information is also provided in the public-use microdata documentation for the year of interest. (See, for example, "2016 Users' documentation, Interview Survey, Public-Use Microdata (PUMD), Consumer Expenditure," August 29, 2017, https://www.bls.gov/cex/pumd.htm.)

8 For example, suppose a respondent reports values for two sources of income: (1) wages and salaries and (2) pensions. Further suppose the following: The reported value for wages and salaries exceeds the critical value, and is therefore replaced by the topcoded value of $X; the reported value for pension income, $Y, is below the critical value for this income source; and the value for total income is shown to be $X + $Y + $Z. Because this respondent only has two sources of income reported and pension income is not topcoded, the reported value for wages and salaries is $X + $Z. To prevent this, total income must be computed after each individual component has been topcoded as needed. Therefore, in this example, total income is $X + $Y and the actual reported value of wages and salaries cannot be "reverse engineered."

9 The project involved finding and merging results from the FMLI, MEMI, and MTBI files. The FMLI files include general characteristics of the consumer unit (e.g., region of residence, number of members, etc.) and summary variables (e.g., total educational expenditures). The MEMI files contain information on each individual member of the consumer unit (e.g., each member's age, race, educational attainment, etc.). The MTBI files include expenditures for specific educational expenses (e.g., expenditures on "College tuition," "Elementary and high school tuition," "Test preparation, tutoring services," "School books, supplies, equipment for vocational and technical schools," etc.).

10 As expected, in the CE data, income after taxes is simply income before taxes minus taxes paid. However, as with income, the respondent may not know, or refuse to provide, tax data in whole or in part, compounding the problems associated with analyzing income before taxes.

11 For data collected in 2003 and earlier, consumers were classified in income groups based on whether or not they were "complete income reporters," in which generally, the respondent provided values for major sources of income, such as wages and salaries, self-employment income, and Social Security income. However, even complete income reporters may not have provided a full accounting of all income from all sources for all members of the consumer unit. For more details on this topic, see the CE glossary (https://www.bls.gov/cex/csxgloss.htm, accessed August 24, 2018). For information on how publications changed with the implementation of income imputation, see "Description of Income Imputation Beginning with 2004 Data" (https://www.bls.gov/cex/csximpute.htm, accessed August 24, 2018).

12 For details about the process, see “Improving data quality in Consumer Expenditure Survey with TAXSIM,” by Geoffrey Paulin and William Hawk, Monthly Labor Review, March 2015, pp. 1–13 (https://doi.org/10.21916/mlr.2015.5, accessed August 24, 2018).

13 Because the practical training is progressive, until 2011 this activity was held on the second day to maximize overlap in attendance between newer and more experienced users. However, in response to comments from attendees at prior workshops, in 2012 the activity was scheduled for the first day of the workshop and successfully repeated in this order subsequently.

14 Specifically, attendees learned how to access the EDA files to ascertain for what type of school or facility (college or university, elementary through high school, child day care center, etc.) certain educational expenditures were incurred, and whether the expenditures were for a member of the consumer unit or a gift to someone outside of it.

15 For example, suppose the sample consists of two consumer units, one of which represents 10,000 consumer units in the population (i.e., itself and 9,999 others like it) and another that represents 20,000 consumer units in the population. If the first spent $150 and the second spent nothing (i.e., $0), the sample mean expenditure is $75. But the population-weighted mean is $50, or [($150 x 10,000)+($0 x 20,000)]/(10,000 + 20,000).

16 In the CE, the term “food at home” generally refers to the location of purchase, not place of consumption, of the food. That is, according to the CE glossary, “Food at home refers to the total expenditures for food at grocery stores (or other food stores) ….” (https://www.bls.gov/cex/csxgloss.htm) Food purchased from restaurants, food trucks, vending machines, etc., are considered to be “food away from home,” even if they were taken home and eaten there.

17 As noted in the introduction to the workshop, the Interview Survey collects data from respondents for four consecutive 3-month periods. During each interview, the respondent is asked to provide information on expenditures for various items during the previous 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 remain. Therefore, attempting to analyze average annual expenditures by only examining respondents who participate for all four interviews yields biased results.

18 For example, normalizing the data can reduce heteroscedasticity in a regression framework.

19 These transformations are often known as “Box-Cox” transformations, after authors G.E.P. Box and D. R. Cox, who wrote a seminal paper about them. (“An Analysis of Transformations,” Journal of the Royal Statistical Society, Series B (Methodological), Vol. 26, No. 2 (1964), pp. 211–252.)

20 At present, no consensus has emerged on a name for this group. The CE program has previously followed the nomenclature of the Pew Research Center, which officially defined this group as “post-Millennials” on March 1, 2018. For more information, see “Fun Facts about Millennials: comparing expenditure patterns from the latest through the Greatest generation,” Monthly Labor Review, March 2018, https://doi.org/10.21916/mlr.2018.9, esp. endnote 14; and a Pew Research Center report by Michael Dimock, “Defining generations: Where Millennials end and post-Millennials begin,” (http://www.pewresearch.org/fact-tank/2018/03/01/defining-generations-where-millennials-end-and-post-millennials-begin/).

21 Attendees also appreciated the information provided in advance of the workshop, which helped in planning their travel, entry to the building, and anticipating what to expect while attending. For these and other details, see https://www.bls.gov/cex/information-for-2019-workshop-attendees.pdf.

RELATED CONTENT

Related Articles

Consumer Expenditure Surveys Methods Symposium and Microdata Users’ Workshop, July 18–21, 2017, *Monthly Labor Review*, June 2018.

Consumer Expenditure Surveys Methods Symposium and Microdata Users' Workshop, July 12–15, 2016, *Monthly Labor Review*, May 2017.

Consumer Expenditure Survey Methods Symposium and Microdata Users' Workshop, 2015, *Monthly Labor Review*, June 2016.

Related Subjects

Survey methods | Sampling | Consumer expenditures | Survey procedures