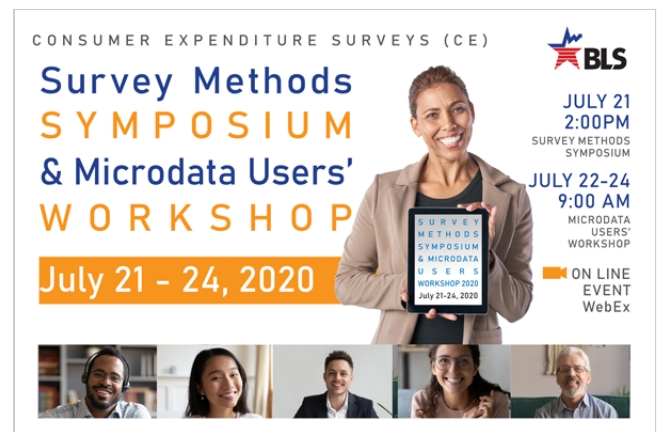


# Consumer Expenditure Survey Methods Symposium and Microdata Users' Workshop, July 21–24, 2020

*The Consumer Expenditure Surveys (CE) program collects expenditure, demographic, and income data from families and households. The CE program held its annual Survey Methods Symposium and Microdata Users' Workshop from July 21 to 24, 2020, 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. 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 virtually to explore better ways to collect CE data and to learn how to use the microdata once they are produced. The experience was unique for presenters and attendees alike in that this was the first time either event was held online, in whole or in part.*

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”).<sup>1</sup> 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).<sup>2</sup> Researchers use these data in a variety of fields, including academia, government, and various private industry areas, such as market research.

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 in which the data are used and thus provide opportunities to explore collaboration. In 2009, the



### Geoffrey D. Paulin

[paulin.geoffrey@bls.gov](mailto: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](mailto:krishnamurty.parvati@bls.gov)

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

workshop expanded from 2 days to 3 days to include presentations from data users not affiliated with BLS. This expansion 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.

In every year from 2012 onward, a 1-day symposium has preceded the workshop. The purpose of the symposium is to support 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 2020 Survey Methods Symposium, conducted on July 21, 2020, and the 2020 Microdata Users' Workshop, conducted July 22–24, 2020.

For the first time, in whole or in part, both events were held online, rather than at the BLS national office in Washington, D.C. The CE program made this decision because of the continuing COVID-19 pandemic. As a result, to minimize potential disruption due to possible technological failures or other, unanticipated, problems, both events were streamlined. For example, the Symposium, which usually features several speakers outside the CE program, instead consisted of only two presentations, both from CE staff. While the workshop maintained its tradition of featuring outside (non-CE program) speakers, BLS speakers, which usually include staff from several programs, were limited to members of one branch (Information and Analysis) of the CE program. The one exception was an overview of CE data presented by the CE program director.

## Survey methods symposium

The symposium began with a presentation on the Gemini redesign titled “Gemini Redesign: Past, Present, and Future” by Parvati Krishnamurty from the CE program at BLS. The presentation outlined the original plans for the redesign and recent modifications made to the redesign plan for implementation. The redesign plan was intended to be implemented as a whole, but because of budget constraints, the plan will instead be implemented in phases. Therefore, the plan was modified to move to a phased implementation of key design elements into the CE surveys. The phased implementation plan is to retain the design elements that have been effective during field tests, which include a streamlined questionnaire with less expenditure detail, records focus (including a targeted incentive for record use), online diaries, and token incentives.<sup>3</sup> These elements will 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 could be tested and implemented in future years, pending changes to requirements or funding availability. Dr. Krishnamurty provided more detail about the Large Scale Online Diary Feasibility Test (LSF), which was fielded from October 2019 to March 2020. She also provided a high-level overview of the streamlined questionnaire design and plans for releasing the new sections of the streamlined questionnaire in three phases starting in 2023. Dr. Krishnamurty mentioned future enhancements that are being explored by the CE, including new technologies such as receipt scanning and geolocation, self-administered interviews, adaptive design, split questionnaire design, single sample design, and gold-standard interviews.

The second presentation by Laura Erhard from the CE program was titled “Going Online: Results from the CE’s LSF.” The presentation summarized the test design, the online diary design, and some preliminary results from the first 3 months of unprocessed data. Procedural changes had to be made to the LSF in March 2020 when the

pandemic made in-person visits impossible, but otherwise fielding went smoothly. The overall response rate was 47.2 percent, and the rate of online diary placement was lower than expected, despite screening respondents for internet access and frequency of use. Barriers to online diary placement reported by field representatives include language issues, lack of technological savviness, and lack of connectivity. The LSF included two experiments: an advance postcard and a token incentive. While there was a small and nonsignificant increase in response rates from a \$5 token incentive, there was a large but nonsignificant increase in response rates from advance postcards in the preliminary data.<sup>4</sup> In general, respondents reported positive experiences with the online diary. One area of concern was the large number of failed respondent logins to the online diary. Although the online diary has been used as a contingency measure in the CE Diary Survey during the COVID-19 pandemic, the CE staff is conducting data analysis of the online diary and will consider its use for production in 2022. Additionally, CE and Census staff are working on making improvements to the online diary design, protocols, and training based on lessons learned from the LSF.

## Microdata users' workshop

*Meet with an expert:* Beginning with the 2017 workshop, the CE organizers have included a feature called the "Meet with an expert" program. The purpose of the program is to provide an opportunity for attendees to have in-depth, one-on-one meetings with members of the CE staff, during which the attendees can ask questions and receive comments and other guidance about the projects in which they are engaged.<sup>5</sup> With the workshop shifting online this year, the meetings were reimagined. Instead of conducting meetings, attendees met with their experts by phone, calling a prearranged toll-free number at their appointed times. In addition, several of those who were waitlisted because of unusually high demand scheduled their meetings for the week following the workshop.

The program has proven beneficial to attendees and to CE staff, who learn more about how researchers are using the data and about factors related to data, documentation, etc., that can be improved. Despite the differences in the mode of meeting (i.e., by phone instead of in person), the program was just as successful at the 2020 workshop. During the feedback session, those who participated in this program unanimously praised the experience both for the content of the meeting and the quality of information received. As a result, the program will be continued for the 2021 Microdata Users' Workshop. Attendees are able (and encouraged) to arrange meetings via the registration form or email.

## Day one

The first session of the 2020 workshop consisted of presenters from the CE program. After welcoming remarks by Scott Curtin, chief of the Branch of Information and Analysis (BIA), Adam Safir provided an overview of the CE, featuring topics including how the data are collected and published. Economist Bryan Rigg (BIA) 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. Mr. Curtin completed the session with a description of the data file structure and variable naming conventions.

Afterward, attendees received their first practical training with the data. In this session, led by senior economist Aaron Cobet (BIA), 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).<sup>6</sup> As expected, the circumstances of the workshop offered new challenges for this training session and subsequent training sessions: When conducted in person, members of CE staff circulate among attendees to offer help. In addition, some

attendees choose to work together on the projects. While these features were obviously not available for the online workshop, attendees were able to submit questions via an online chat feature or send email to CE staff to receive an answer directly or arrange a phone call with CE staff.

The afternoon activities included presentations from researchers not affiliated with the CE program. Summaries of the papers presented by outside researchers are included at the end of this report.

The first speaker, data scientist Aaron R. Williams (Urban Institute, Income and Benefits Policy Center), spoke about his use of CE microdata to study income and expenditures by low-income families with at least one member age 50 or older. The work was coauthored with Damir Cosic (Senior Research Associate, Urban Institute, Income and Benefits Policy Center), who attended the 2019 workshop.<sup>7</sup>

The second presentation was codelivered by Casey Goldvale (policy analyst) and Vincent Palacios (senior policy analyst) of the Georgetown Center on Poverty and Inequality. They described their work investigating costs beyond tuition (e.g., housing) for older (age 25 to 45) college students.

Following this presentation, self-directed practical training resumed with projects, introduced by economist Jimmy Choi (BIA), involving the integration of data from the Diary and Interview Surveys, a practice used in production of CE tables, and finding detailed information about education expenditures.<sup>8</sup> Attendees also learned how to integrate results from the Interview and Diary Surveys to match expenditure categories in CE published tables.

After this session, the workshop concluded for the day.<sup>9</sup>

## Day two

To open the second day, Mr. Cobet explained the need to balance confidentiality concerns of respondents with the usefulness of the data to researchers. Because U.S. Code Title 13 requires confidentiality of response, information that might 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 to produce the CE microdata files to address these disclosure concerns. The first method, called topcoding, involves reported values for income or expenditures that exceed a certain threshold, called the critical value. These top-coded values are replaced by an average of all values exceeding this critical-value threshold and then flagged as topcoded (or bottom-coded, in the case of large income losses).<sup>10</sup> 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 to a nearby state) 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).<sup>11</sup> 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.<sup>12</sup>

Next, Mr. Choi presented a brief description of experimental weights for estimating state-level expenditures with the use of the CE microdata. He noted that weights for New Jersey, California, Florida, New York, and Texas were available (<https://www.bls.gov/cex/csresearchtables.htm#stateweights>).<sup>13</sup> Mr. Choi also presented the criteria

used by the CE program to assess the feasibility of devising weights for other states (sample size, confidentiality concerns, and long-term retention of the state under study in the CE sample).

Concluding the session, Dr. Geoffrey Paulin, senior economist in the CE program (BIA), described the correct use of sample weights in computing consumer unit population estimates. His talk started with an overview of the computation of the weights.<sup>14</sup> Following this, he introduced the procedures needed to get 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.<sup>15</sup> Finally, he noted that the proper use of weights requires a special technique, called balanced repeated replication (BRR). BRR accounts for sample design effects in order to produce correct estimates of variances for weighted means, regression parameters, etc. Without BRR, these estimates can be biased or otherwise incorrect when computed for CE data. Next, he provided an example of BRR he derived from a question that arose during the talk. This led into a practical training session, instructed by Mr. Curtin, 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 the expenditures actually occurred. For example, for a person participating in the Interview Survey in January 2018 who reports expenditures that occurred during the final 3 months of 2017 (i.e., October, November, or December), the expenditure collection year is 2018, while the expenditure calendar year is 2017.

Presentations by non-CE staff researchers continued in a themed session during the afternoon. Each of the speakers described their work with race and ethnicity variables in the CE microdata. The first presenter, Ziyao Tian, a Ph.D. candidate in sociology at Princeton University, discussed expenditures on higher education for Asian-American families. The second, copresented by Reginald Noël (research economist/data scientist) and Whitney Hewitt Noël (public health researcher/health equity advocate), both of the Noël Collective, discussed the intersectionality of sex and race in both income and expenditure patterns. Serving as a moderator of the discussion, Dr. Paulin briefly described his own work with the Diary Survey to explore food expenditures by race and ethnicity. He noted the detailed information on geographic origin included in race (Asian) and ethnicity (Hispanic) categories within the CE data for users interested in studying expenditures by the communities within these broader groups.<sup>16</sup> He also pointed out the benefit of having these characteristics available for each member of the consumer unit, which he applied to his own research. For example, there is no attempt to identify a “decision maker” in the consumer unit, so the relationship of race or ethnicity to expenditure patterns is unclear when members of the consumer unit identify with different races or ethnicities.<sup>17</sup>

The last session of the day continued practical training. Dr. Paulin described the proper methods for analyzing CE income data, which are multiply imputed when missing. This presentation led into more self-directed practical training, in which attendees applied the methods described.

## Day three

The final day started with a set of presentations from outside researchers who use CE microdata. The first of these presenters was Dr. Constantin Burgi, professor of economics at St. Mary's College of Maryland. Dr. Burgi discussed his work examining how average consumer expectations differ when reporting households are weighted by actual expenditures, as opposed to households having equal weight, in computing the average.

The second speaker, Dr. Ensieh Shojaeddini, a researcher on fellowship at the Environmental Protection Agency, used CE data in the construction of demand systems to estimate effects of regulation.

The final speakers in this session were Dr. David King, an assistant professor of urban planning at Arizona State University, Tempe, and Dr. Jonathan Peters, a professor of finance and data analytics at The City University of New York. The presenters noted several changes in the last decade that affect transportation expenditures for consumers (the rise of rideshare services, online shopping, and, most recently, the COVID-19 pandemic), and want to see how these changes will continue to affect these expenditures in the future. They proposed a plan for studying patterns using CE data and other sources, particularly once the CE data for 2020 are released.

Following a break, 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. Next, supervisory economist Brett Creech (BIA) provided a sneak peek of developments for CE publications and microdata. Starting with those recently implemented, such as the release of free PUMD covering 1980 to 1995 in early 2020,<sup>18</sup> he described changes scheduled or under consideration for future releases.<sup>19</sup> Those releases scheduled include new tables showing expenditures in 2019 at more refined geographic levels (census division in addition to current census region) and a new column on the generational tables (first published officially to reflect 2016 data) showing expenditures for the post-Millennial generation (i.e., those born in 1997 or later).<sup>20</sup> He also noted the addition of a new question (July 2018 for the Interview Survey and January 2019 for the Diary Survey) that asks whether anyone in the consumer unit has previously served in the U.S. military. He stated that tables showing expenditures by veteran status will be published as soon as sample size permits. In addition, he announced the inclusion of a special question, starting in June 2020, regarding the receipt and use of the 2020 economic stimulus payments. Both microdata and published tables will include information collected from the special question.<sup>21</sup>

To conclude the workshop, David Biagas (BLS) led attendees in a feedback session. During the feedback 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. Many comments were positive, with attendees liking the progressive nature of the workshop (i.e., starting with the most basic information about the data collection and file structures and ending with the most technical topics) and praising the “Meet with an expert” program. Workshop attendees also provided suggestions on what could be improved. These comments were especially important given the delivery of the workshop online this year, for the first time ever. Because of the ongoing COVID-19 pandemic, the workshop will be conducted online again in 2021.

## Symposium and workshop of 2021

The next Survey Methods Symposium is scheduled for July 20, 2021, in conjunction with the 16th annual Microdata Users’ Workshop (July 21–23). Both events will be held online. Although the symposium and workshop remain free of charge to all participants, advance registration is required (<https://data.bls.gov/cgi-bin/forms/cex-registration>). For more information about these and previous events, visit the CE website (<https://www.bls.gov/cex/>) and look for the left navigation bar, titled “CE WORKSHOP AND SYMPOSIUM.” For direct access to this information, the link is <https://www.bls.gov/cex/csxannualworkshop.htm>. The link to the combined agendas for the 2020 symposium and workshop (<https://www.bls.gov/cex/ce-2020-combined->

[agenda.pdf](#)) is also available on this webpage. Workshop presentations are available in an online zip file (<https://www.bls.gov/cex/ws2020-presentations.zip>).

## 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.

Aaron R. Williams, Data Scientist, Income and Benefits Policy Center (Urban Institute), “Lifetime Income & Costs of the LI50+” (Interview Survey), day one.

A primary mission of the AARP Foundation is to mitigate, and eventually eliminate, poverty among older Americans. An important concern for the Foundation in addressing poverty among seniors is to select the target population that maximizes the impact of their effort. Our report, which focused on households below 250 percent of the Federal Poverty Guideline with at least one member age 50 or older, helped the AARP Foundation 1) identify demographic groups that are most vulnerable and 2) identify the groups that represent the biggest share of the vulnerable population. To identify the most vulnerable population—those in high need—we relied on household spending rather than income because it is measured more accurately than income and represents a better measure of personal well-being. We selected the bottom expenditure quartile—25 percent of LI50+ who had the lowest annual expenditures adjusted for household size—and analyzed the composition of this group and the likelihood of being in high need among the general population. Through this work, we developed a customized version of the R package library (cepumd) by Arcenis Rojas, we created a custom mapping of Universal Classification Codes to a custom hierarchy of grouped expenditures that matched the interests and needs of the AARP Foundation, and we created a detailed profile of the consumption and income of the LI50+ with extensive data visualization and tables.

We used a heavily functional approach in R to analyze the data and built a process with version control that proved useful for this analysis and hopefully future analyses.

Casey Goldvale, Policy Analyst, and Vincent Palacios, Senior Policy Analyst, Georgetown Center on Poverty & Inequality, “Costs Beyond Tuition: Estimating older college students’ basic needs with Consumer Expenditure Survey (PUMD) (Interview Survey), day one.

Though estimating the “cost of attendance” is key in determining student financial aid for higher education, there are no standardized measurement methods and estimates can vary wildly across colleges located within a few miles of each other. There is also evidence that “cost of attendance” may be severely underestimated for older students who are more likely to have dependents and be financially independent. We use the Consumer Expenditure Surveys (CE) to estimate average spending on components of an adequate living standard among older undergraduate students’ households nationally. Using UCC codes from MTBI data files, we have adapted FMLI and MEMI samples and variables to be comparable to cost categories defined in U.S. student financial aid policy and the Census Bureau and BLS basic needs and

poverty measurement methodologies. We also focus on equity by incorporating race/ethnicity, gender, and other demographic and geographic characteristics for older students and their households. To ensure adequate sample sizes, we pooled multiple years of data to increase sample size and adjusted the sampling weights accordingly. To our knowledge, this is the first time the CE has been used to study the older student population and is one of few studies beyond Geoffrey Paulin's 2001 paper using CE microdata to estimate college students' cost of living.<sup>22</sup>

Ziyao Tian, Ph.D. Candidate (Sociology), Princeton University, "How Expensive Is the Battle of Tiger Mothers? Understanding Race and Class behind the Educational Expenditure of Asian Americans" (Interview Survey), day two.

Social stratification scholars have been trying to understand the superior academic achievement of Asian Americans by examining the roles of family socioeconomic status (SES), culture, and the intersection of the two factors. Yet, the role of expenditure on education as an important mechanism linking social class and culture remains unexplored. Previous studies demonstrate that superior academic achievement is partly driven by Asian Americans' high expectations of education across families of different SES origins. In other words, family SES has a weaker predicting power of educational expectations for Asian Americans than for Whites. Beyond this psychological-attitude channel, we use the Consumer Expenditure Surveys (CE) data from 2009 to 2019 to examine whether Asian Americans' expenditure on education is also universally higher and less sensitive to SES. Preliminary results show that Asian Americans, on average and across SES distribution, spend more dollars, as well as a higher proportion of their spending budget, on education than their non-Hispanic White counterparts. The difference is primarily a result of Asian families' high spending on college tuition. The racial gap in college tuition is more pronounced among lower-SES families than among higher-SES families. Further explorations of the gap in college tuition suggest that the difference is mainly driven by more college students from less advantaged Asian families, rather than a greater tendency to provide stronger college tuition support when having a college student at home.

Reginald Noël, Research Economist/Data Scientist, and Whitney Hewitt Noël, Public Health Researcher/Health Equity Advocate, Noël Collective, "Gender Economics, Race, and Intersectionality: Using CE Microdata to examine inequalities among adult women and men in the U.S. by race and ethnicity, 2016 through 2018 combined" (Interview Survey), day two.

This working paper explores the issues of gender economics, with an intersectional dimension of race and ethnicity. Comparative analysis from two different datasets, the American Community Survey and the CE, show similar persistent inequalities in income, stratified by binary sex and race. Specifically, adult men had higher salaries and wages than adult women. In addition, adult Asian and White populations had higher salaries and wages than the adult Native, Black, and Hispanic populations. Moreover, the Consumer Expenditure Interview Survey data allowed for a deep examination of household spending, scarcity, resource allocation, and consumer patterns among the different cohorts. The data illustrated socioeconomic inequalities faced by women as compared with men, including the Gender Pension Gap, Pink Tax (higher prices for goods marketed to women, such as razors, that are actually or nearly identical to versions marketed to men), health care costs, educational attainment, occupation, and marital status. All



these factors depicted microeconomic inequalities faced by intersected subpopulations, which disservices not only these population groups but also the U.S. economy as a whole.

Constantin Burgi, Ph.D., Assistant Professor (Economics), St. Mary's College of Maryland, "Predicting consumer expenditure based on the variables available in the Consumer Expectation Survey of the NY Fed" (Interview Survey), day three.

The aim of this work is to check how the mean household expectations from the New York Fed's Consumer Expectations Survey change when households are weighted using consumer expenditure, as in the Consumer Price Index, instead of equal weights. In order to do so, it is necessary to impute the consumer expenditure of the households in the Consumer Expectation Survey. Variables that are available in both the CE and the Consumer Expectations Survey are made comparable and a (weighted) OLS regression is then used to impute the consumer expenditure. It is found that the consumption-weighted consumer expectations are around 0.7 percentage points lower than the equally weighted consumer expectations.

Ensieh Shojaeddini, Ph.D., Oak Ridge Institute for Science and Education fellow at U.S. Environmental Protection Agency, "Consumer Demand Estimation for Heterogeneous U.S. Households" (Interview Survey), day three.

The specification of the consumer demand system is important for estimating the economy-wide impacts of environmental regulation. First, it plays a key role in determining the baseline in a dynamic context. Second, it defines the final good demand curves that help determine the ability to control pollution on the extensive margin through the output effect. In this role, the specification of consumer demand also helps determine the share of abatement costs borne by factors or production relative to consumers. Finally, it plays an important role in determining tax interaction effects.

In computable general equilibrium (CGE) models, household behavior is typically governed by a constant elasticity of substitution (CES) utility function, though it fails to realistically capture well-known patterns of consumer behavior. In addition, only a few CGE models econometrically estimate their own elasticities, which are limited to a representative national-level household. We empirically estimate several flexible consumer demand systems for the U.S. economy for use in a CGE model with regional and household income disaggregation. As part of this evaluation, we consider tradeoffs between different specifications regarding complexity, regularity, the ability to capture cross-price elasticities, Engel curve flexibility, and the number of commodities that can be reasonably accommodated.

David A. King, Ph.D., Assistant Professor of Urban Planning, Arizona State University, Tempe, Arizona, and Jonathan Peters, Ph.D., Professor of Finance and Data Analytics, The City University of New York, "Household Transportation Spending Trends from 2010 to 2020 - Early Indications of the impact of cultural shifts and pandemic related household activity on transportation spending patterns" (Interview Survey), day three.

The last 10 years have been a time of radical change in household consumption as it relates to transportation spending. First, we experienced the massive growth in for-hire vehicle services such as Uber and Lyft that disrupted and transformed traditional taxi services in many cities. Second, we observed a decline in private vehicle ownership in several cities, with corresponding growth in car sharing services. Further, the growth in e-bicycles and scooter services, as well as the potential growth for autonomous vehicles, have made the last decade a time of revolution in the transportation sector. Now, further changes

are being wrought by the COVID-19 pandemic, reversing many trends in transportation use. Transit systems reeled from the needs for enhanced sanitation and social distancing, and ridership caps were instituted on many mass transit systems. Demand for gasoline and diesel collapsed. Online shopping and at-home consumption skyrocketed. What is still an open question in all of this is, are these changes temporary and will they reverse when the pandemic moderates, or will they result in a long-term reversal of the recent trends and usher in a 21st century wave of automobile use and reliance on personal instead of shared transportation services? These changes have the potential to disrupt many policy initiatives in terms of infrastructure investment; for example, a shift away from federal funding and a general movement to local funding sources such as tolls or parking fees.

When available, the authors will utilize new data collected on post-COVID-19 consumption from outside sources and compare these sources with BLS CE data to examine how household consumption may have shifted during this period (2010–20). The authors also plan to project what may happen in transportation consumption over the next five years (2021–25).

## Workshop presenters

### Staff of the CE program

Choi, Jimmy. Economist, Branch of Information and Analysis, BIA: practical training leader, day one; presenter, day two.

Cobet, Aaron. Senior Economist, BIA: practical training leader, day one; presenter, day two.

Creech, Brett. Supervisory Economist, Chief, Publications and Tables Production Section, BIA: presenter, day three.

Curtin, Scott. Supervisory Economist, Chief, BIA: emcee, days one, two, and three; practical training leader, day two.

Paulin, Geoffrey. Senior Economist, BIA: introducer of speakers, commentator, days one, two, and three; practical training leader, day two; presenter, days two and three.

Rigg, Bryan. Economist, BIA: presenter, day one.

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

### Other BLS speakers

Biagas, David. Research Psychologist, Office of Survey Methods Research: feedback coordinator, day three.

### Non-BLS speakers

Burgi, Dr. Constantin. Assistant Professor of Economics, St. Mary's College of Maryland, "Predicting consumer expenditure based on the variables available in the Consumer Expectation Survey of the NY Fed" (Interview Survey); day three. First-time attendee and presenter (2020).

Goldvale, Casey. Policy Analyst, Georgetown Center on Poverty & Inequality, "Costs Beyond Tuition: Estimating older college students' basic needs with Consumer Expenditure Survey (PUMD)" (Interview Survey); day one. Former attendee (2019) and first-time presenter (2020).

King, Dr. David (Ph.D.). Assistant Professor of Urban Planning, Arizona State University (Tempe), "Household Transportation Spending Trends from 2010 to 2020 - Early Indications of the impact of cultural shifts and pandemic related household activity on transportation spending patterns" (Interview Survey); day three. First-time attendee and presenter (2020).

Noël, Reginald. Research Economist/Data Scientist, Noël Collective, "Gender Economics, Race, and Intersectionality: Using CE Microdata to examine inequalities among adult women and men in the U.S. by race and ethnicity, 2016 through 2018 combined" (Interview Survey); day two. First-time attendee and presenter (2020).

Noël, Whitney Hewitt. Public Health Researcher/Health Equity Advocate, copresenter with Reginald Noël; day two. First-time attendee and presenter (2020).

Palacios, Vincent. Senior Policy Analyst, Georgetown Center on Poverty & Inequality, copresenter with Casey Goldvale; day one. Former attendee (2019) and first-time presenter (2020).

Peters, Dr. Jonathan (Ph.D.). Professor of Finance and Data Analytics, The City University of New York, copresenter with David King; day three. Prior presenter (2014, and 2017 through 2019); returning presenter (2020).

Shojaeddini, Dr. Ensieh (Ph.D.). Oak Ridge Institute for Science and Education fellow at U.S. Environmental Protection Agency, "Consumer Demand Estimation for Heterogeneous U.S. Households" (Interview Survey); day three. Former attendee (2019) and first-time presenter (2020).

Tian, Ziyao. Ph.D. Candidate (Sociology), Princeton University, "How Expensive Is the Battle of Tiger Mothers? Understanding Race and Class behind the Educational Expenditure of Asian Americans" (Interview Survey); day two. First-time attendee and presenter (2020).

Williams, Aaron R. Data Scientist, Income and Benefits Policy Center (Urban Institute), "Lifetime Income & Costs of the LI50+" (Interview Survey); day one. First-time attendee and presenter (2020).

SUGGESTED CITATION

Geoffrey D. Paulin and Parvati Krishnamurty, "Consumer Expenditure Survey Methods Symposium and Microdata Users' Workshop, July 21–24, 2020," *Monthly Labor Review*, U.S. Bureau of Labor Statistics, June 2021, <https://doi.org/10.21916/mlr.2021.14>

## NOTES

<sup>1</sup> 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/glossary/#term\\_Household](https://www.census.gov/glossary/#term_Household) and [https://www.census.gov/glossary/#term\\_Familyhousehold](https://www.census.gov/glossary/#term_Familyhousehold).

<sup>2</sup> 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](https://www.bls.gov/cex/pumd_data.htm).

Data from the Diary and Interview Surveys are published twice a year in various standard tables. One set describes expenditures that occurred within the calendar year of interest (e.g., January through December 2018 for the most recent set available as of the writing of this report). The other set provides a midyear update to expenditures, ranging from July of the earlier year to June of the later year (e.g., July 2017 through June 2018 for the most recent set available as of the writing of this report). The single-year series is available from 1984 onward. The midyear updates are available from July 2011 to June 2012 onward. Each set includes information on expenditures by age of reference person, composition of consumer unit, income of consumer unit, and other demographics. For a complete list, see <https://www.bls.gov/cex/tables.htm>.

<sup>3</sup> Token incentives are being tested in the LSF prior to potential implementation in the CE.

<sup>4</sup> Since the symposium, analysis of LSF data from October through February indicates that postcards have no impact on response rates.

<sup>5</sup> Attendees were able to sign up for a meeting by checking a box on their registration forms. They could also sign up via e-mail throughout the virtual workshop, replacing the option to do so at the registration desk for previous in-person workshops previously. 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.

<sup>6</sup> The projects in this series built on each other, progressing from basic computation to more complicated use of the data, which 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.).

<sup>7</sup> In an email exchange with the author of this workshop report, Dr. Cosic states, “...my attendance of the 2019 workshop was instrumental in our successful completion of the project that Aaron presented. I think this is an excellent example of the success of your workshop.” (email from Damir Cosic to Geoffrey Paulin, August 2, 2020).

<sup>8</sup> 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 daycare center, etc.) certain education expenditures were incurred, and whether the expenditures were for a member of the consumer unit or a gift to someone outside of it.

<sup>9</sup> From 2012 until 2019, the first day of the workshop ended with a networking opportunity, where attendees could meet each other and informally discuss questions with CE staff. (Prior to 2012, this event was held on the second day of the workshop, to maximize overlap in attendance between newer and more experienced users.) However, with the delivery of the workshop online, this activity was unfeasible, due to limitations of the software approved for delivering the workshop.

<sup>10</sup> 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.

<sup>11</sup> For details on topcoding and suppression, including specific variables affected and their critical values, see [https://www.bls.gov/cex/pumd\\_disclosure.htm#Basic](https://www.bls.gov/cex/pumd_disclosure.htm#Basic).

<sup>12</sup> 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.”

<sup>13</sup> Weights for the first three states (New Jersey, California, and Florida) are available for 2016 onward; for the latter two (New York and Texas), they are available for 2017 onward.

<sup>14</sup> Traditionally, preceding this talk, a member of the Statistical Methods Division delivers a detailed explanation of the computation of the weights. However, as noted earlier, the workshop planners cut several detailed presentations due to the uncertainties of the first-ever online workshop.

<sup>15</sup> 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. However, the population-weighted mean is \$50, or  $[(\$150 \times 10,000) + (\$0 \times 20,000)] / (10,000 + 20,000)$ .

<sup>16</sup> That is, in addition to asking the respondent the race of each member of the consumer unit, if Asian, the Interview and Diary Surveys both ask about geographic origin: Chinese, Filipino, Japanese, Korean, Vietnamese, Asian Indian, or other Asian (listed in the order of appearance in the questionnaire). If the respondent reports that a member is Hispanic, the interviewer asks whether the member is Mexican, Mexican American, Chicano, Puerto Rican, Cuban, or other Spanish (again, listed in the order of appearance in the questionnaire).

<sup>17</sup> Even if a question were asked about “decision making,” it is not clear that the answer would be meaningful in a “real world” context. For example, in married couples, it is likely that at least some decisions are made jointly, and in those that are not, it is not clear who makes the decisions. For example, if only one spouse purchases the groceries, which spouse is it? Furthermore, that spouse will almost certainly take into account preferences of the other spouse. If the purchasing pattern therefore reflects the tastes (literally) of both spouses in food consumption, and these tastes are influenced by the different racial or ethnic backgrounds of each spouse, then the relationship of expenditure to race or ethnicity is diluted within such families. Therefore, comparing consumer units in which all members share the same race and ethnicity makes the comparisons across racial and ethnic groups much clearer.

<sup>18</sup> Available at [https://www.bls.gov/cex/pumd\\_data.htm](https://www.bls.gov/cex/pumd_data.htm).

[19](#) Prior to February 2020, free PUMD were available from 1996 onward. The release of data from 1980 to 1995 allows users to obtain these data for all years in which CE data were collected on a continual basis. Prior to 1980, they were collected approximately every 10 years (1972–73, 1960–61, etc.).

[20](#) While a table showing expenditures for the “post-Millennial” generation for July 2018 through June 2019 is available, the 2019 table will be the first standard (i.e., calendar year) table published to feature expenditures for this group.

[21](#) This question is predated by similar questions added regarding earlier stimulus payments. The first was added in response to payments made in 2001; the second was added in response to payments made in 2008. (See <https://www.bls.gov/cex/anthology11/csxanth5.pdf>.) In addition, CE collected information on the special \$250 payment made in 2009 to most Social Security recipients and other eligible persons. (See <https://www.bls.gov/opub/btn/archive/how-consumers-used-the-2009-economic-recovery-payments-of-250.pdf>.)

[22](#) See “Expenditures of college-age students and nonstudents,” *Monthly Labor Review*, July 2001, pp. 46–50, <https://www.bls.gov/opub/mlr/2001/07/ressum1.pdf>.

RELATED CONTENT

**Related Articles**

[Consumer Expenditure Survey Methods Symposium and Microdata Users’ Workshop](#), July 16–19, 2019, *Monthly Labor Review*, April 2020.

[Consumer Expenditure Survey Methods Symposium and Microdata Users’ Workshop](#), July 17–20, 2018, *Monthly Labor Review*, May 2019.

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

**Related Subjects**

[Survey methods](#) | [Expenditures](#) | [Sampling](#) | [Consumer expenditures](#) | [Survey procedures](#)