

Source Selection for CE PUMD and Publication Tables

Barry Steinberg / Joshua Klick / Bryan Rigg
Consumer Expenditure Surveys / Consumer
Price Index

2023 Microdata Users Workshop
July 18, 2023



What is Source Selection

- Methods used to select the appropriate survey (Interview/Diary) for publication table estimates.
- For items that are unique to one survey or another, the choice is obvious.
- For items that overlap in coverage between the surveys, source selection methods are used to determine which source to select for publishing the Integrated data.



Purpose

- Primarily used for publication tables.
 - ▶ Identifies the more reliable source of survey data to use in estimation.
- PUMD data users
 - ▶ Provides a means for users to integrate survey estimates and closely replicate the publication tables.
- CPI
 - ▶ Monthly expenditure weights used for Chained-CPI-U
 - ▶ Annual Expenditures for the CPI-U



Methodology – Overall Goal

- Over 200 UCCs are processed using both the Diary and the Interview.
- A determination is made to which source is used for the integrated tables.
- The overall methodology selects the higher mean given two decision criteria with exceptions from the CPI.



Methodology

■ Preliminary steps:

- ▶ Data are top coded at 99th percentile and bottom coded at 1st percentile
 - This is done to minimize the impact of outliers
- ▶ Calculating counts, sample means, and sample variances.

Decision Criteria

- There are 2 criteria and a CPI Seasonality consideration that are used in determining source selection decisions:
 - ▶ Criterion 1: Counts Sufficiency
 - ▶ Criterion 2: Statistical Significance (Z-scores)

Criterion 1: Counts Sufficiency

- A sufficient amount of data exists when the count for each of the 3 years is greater than or equal to 60.
- If both surveys have sufficient data, then proceed to the next Criterion (Z-score).
- **If both surveys lack sufficient data, then keep existing source.**

Criterion 2: Statistical Significance (Z-scores)

- If the value of the weighted Z-Score

$$z^* = (3/6)z_{t-1} + (2/6)z_{t-2} + (1/6)z_{t-3}$$

- ▶ Greater than 1.645 – Interview Survey
- ▶ Less than -1.645 – Diary Survey

Frequency

- Process is **now** conducted **every year**
 - ▶ List of UCCs is provided by a BLS production job
 - ▶ Next review is in Fall of 2023 to examine source changes to be put in for 2023 data which will be published in 2024



CE-CPI Review

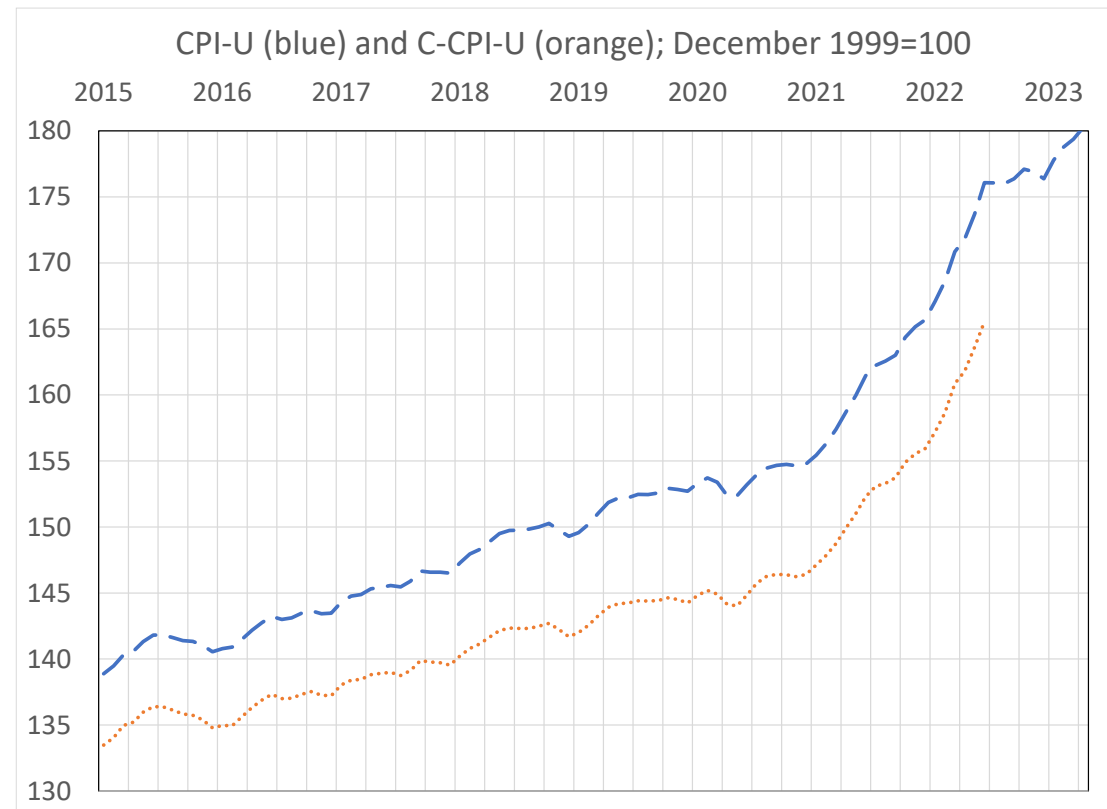
- CE and CPI review the list of UCCs and make a joint decision on which UCCs should change sources
- Reasons to keep existing sources include preventing frequent source selection changes and other considerations
- In addition, there is analytical judgement involved in decisions
- Brett Creech and Barry Steinberg: CE Source Selection for Publication Tables

<https://www.bls.gov/cex/anthology11/csxanth3.pdf>



Chained CPI compared to CPI-U

- **C-CPI-U: monthly weights reflect substitution across items**
 - ▶ lags CPI-U by 10-12 months
 - ▶ Annualized growth rate of 2.3% (2000)
- **CPI-U: fixed quantity weights**
 - ▶ Published 10 days following month
 - ▶ Chained annually beginning 2023
 - ▶ Annualized growth rate of 2.5% (2000)



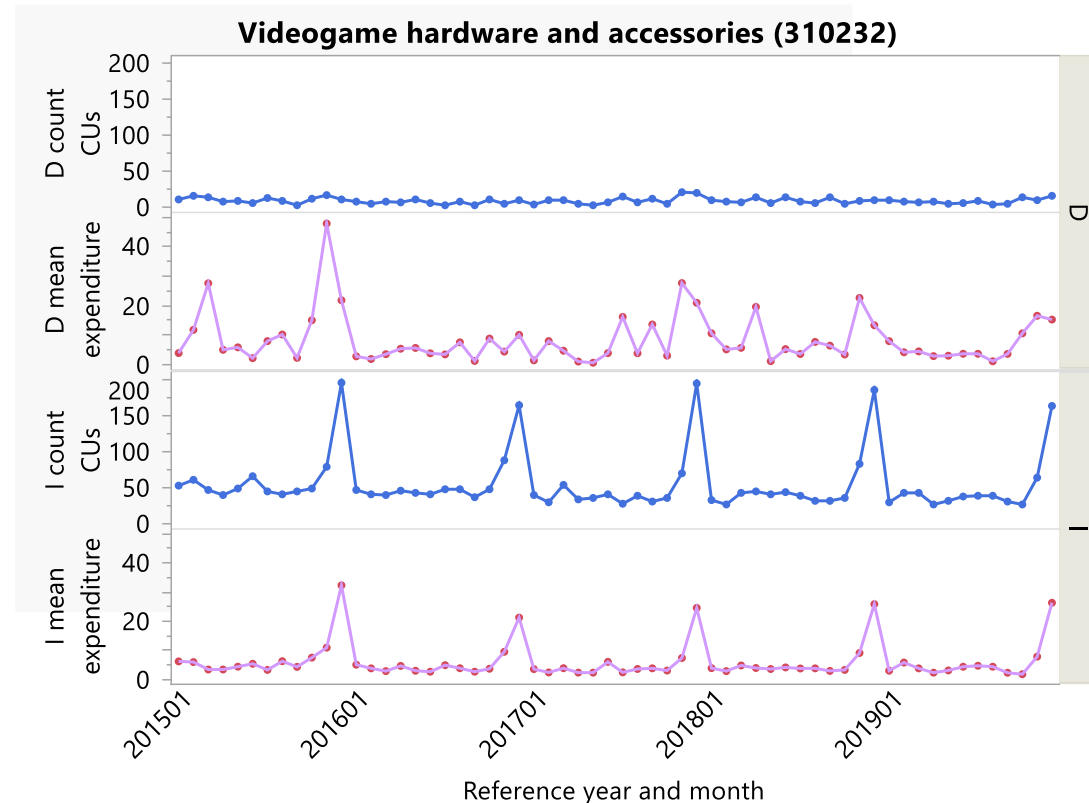
Chained CPI seasonality analysis-monthly weights

■ CPI source analysis-

- ▶ Identify seasonal overlap candidates
- ▶ Autocorrelation (ACF) analysis
- ▶ Compare to secondary sources (ex. PCE)
- ▶ Submit to source selection team

■ Analysis of Videogames (310232):

- ▶ Diary exhibits mean December peaks for 3 of 5 years with high monthly variation
- ▶ Interview exhibits December peaks with low variation across months



Current	CPI ELI	D lags significant	I lags significant
D	RE012	No months	Month 12

What's next?

- 2023 analysis: Continue use of current methodology + CPI contribution
- 2023 data: Number of Overlap UCCS to drop with additional consolidations
- 2024 data: Edit rate criterion to be evaluated



Source selection access

https://www.bls.gov/cex/pumd_doc.htm



U.S. BUREAU OF LABOR STATISTICS

Follow Us | Release Calendar | Blog

Search BLS.gov

HOME | SUBJECTS | DATA TOOLS | PUBLICATIONS | ECONOMIC RELEASES | CLASSROOM | BETA

Bureau of Labor Statistics > Consumer Expenditure Survey > Methods

Consumer Expenditure Surveys

Search Consumer Expenditure Surveys Go

CE Home | CE Publications | CE Data | CE Methods | About CE | Contact CE

PUMD Documentation

This page contains the documentation for the CE public-use microdata (PUMD).

All year-specific Interview and Diary Survey documentation has been combined.

- [PUMD Getting Started Guide](#) provides documentation for the CE PUMD, its conventions, files, and methodology.
- [Dictionary for Interview and Diary Surveys \(XLSX\)](#) provides variables and codes.
- [Sample code](#) provides code in SAS, STATA, and R to prepare annual averages using the data in the Interview Survey, Diary Survey, or both surveys (called "Integrated") and the results. The code may need to be adjusted for specific years.

Additional resources

- [Survey materials](#) lists the documents used to collect data for the Interview Survey and the Diary Survey.
- [Hierarchical groupings \(zip\)](#) includes a description of each UCC along with its hierarchical standing within each expenditure or income category for a given year. Hierarchical groupings are not available for years prior to 1996. For years 1996 and forward, three file types are available:
 - [Integrated groupings](#) lists UCCs that the [CE tables](#) use, and identifies the survey source for the UCCs. These files use this naming convention: CE-HG-Integ-2017.
 - [Interview groupings](#) lists the UCCs from the Interview Survey. These files use this naming convention: CE-HG-Inter-2017. Not available for 1996.
 - [Diary groupings](#) list the UCCs from the Diary Survey. These files use this naming convention: CE-HG-Diary-2017. Not available for 1996.For each UCC, the files provide this information:

Column	Content	Start position	Options
1	Type of information in the line	1	1: First line of name 2: Second line of name *: Comment line
2	Level of aggregation	4	Highest level of aggregation is 1 and lowest 9.

• [Source selection file \(XLSX\)](#) identifies the survey source for the characteristics, expenditure, and income items in the [CE tables](#). Information on source selection is available starting with 1996.



Source Selection reference 1996-2022

1	Survey Source of Data for Consumer Expenditure Survey Integrated Tables																	
2	<i>The detailed list of characteristics, income, and expenditure items below shows which component—the Diary Survey ("D") or the Interview Survey ("I")—was used as the source for that item in the published Consumer Expenditure Survey data tables for each year</i>																	
3																		
4	Level	Description	UCC	y22	y21	y20	y19	y18	y17	y16	y15	y14	y13	y12	y11	y10	y09	y08
5	1	Integrated stub parameter file	HEADINTG	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H
6	1	Number of consumer units (in thousands)	CONSUNIT	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S
7	1	Lower limit	QUINTLIM	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S
8	1	Percent distribution of consumer units	CUDISTRB	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S
9	1	Number of sample diaries	SAMPDIAR												S			
10	1	Consumer unit characteristics:	TITLECU	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
11	2	Income before taxes	INCBFTAX	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
12	3	Meals as pay	800700	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I



4	Level	Description	UCC	y22	y21	y20	y19	y18	y17	y16	y15	y14	y13	y12	y11	y10
897	4	Video game hardware and accessories (new UCC Q20111)	310232	I	I	D	D	D	D	D	D	D	D	D	D	



Contact Information

Barry Steinberg

Steinberg.Barry@bls.gov

Joshua Klick

Klick.Joshua@bls.gov

Bryan Rigg

Rigg.Bryan@bls.gov

