**Using Nontraditional Data Sources to Reduce Respondent Burden in United States Census Bureau Economic Data Products**

Rebecca J. Hutchinson, Economic Directorate, United States Census Bureau of the Census
rebecca.j.hutchinson@census.gov

**QUESTION** Increased respondent burden has led to declining response rates for many of the United States Census Bureau's economic data products, including the indicator programs. Many high-burden and/or non-reporting companies already provide comparable data to private sector companies for market research purposes. Can that same data be used by the Census Bureau to reduce respondent burden while maintaining or enhancing the quality of its published data? Building upon a successful proof-of-concept project that demonstrated promise that retail point-of-sale data can be used in this manner, the project has been expanded to answer the following questions:

- Can national-level tabulations from point-of-sale data be used in place of reported or imputed retailer data in the Monthly Retail Trade Survey and the Annual Retail Trade Survey? Can store-level tabulations from point-of-sale data be used in place of reported or imputed retailer data in the Economic Census? Which quality metrics will be used to make the determination that the point-of-sale data are of good quality?
- Can a mapping of a point-of-sale data product dictionary to the North America Product Classification System (NAPCS) produce product data tabulations that are either comparable or an improvement to what retailers reported to the 2017 Economic Census?
- Can a streamlined IT and survey processing framework be established to allow for an automated, seamless transfer of the data and efficient quality review process?

**DATA** Aggregated point-of-sale datasets purchased from The NPD Group, Inc. contained monthly data for individual retailers by store and product level, i.e. for a given month, sales for Product Z in Store Location Y for Retailer X. The Census Bureau provided dataset requirements to NPD and NPD curated the datasets from their data feeds. The datasets were limited to stores located in the United States and included values for the following variables: time period (month/year), retailer name, store number, zip code of the store location, channel type (brick & mortar or e-commerce), imputation flag, product classifications by industry, category, and subcategory, and sales figures. One observation for each month/year for each store includes a total sales value of the unclassified data which contains all products that NPD does not currently classify.

**METHOD** During the proof-of-concept effort, national-level monthly whole-store tabulations (brick & mortar sales plus e-commerce sales) between the NPD data and the MRTS data, store-level locations and sales between NPD data and 2012 Economic Census, and product-line categories in the NPD data and in the 2012 Economic Census were compared. In the expanded effort, the same comparisons will be made along with comparisons to newly reported 2017 Economic Census sales by store location and by NAPCS.

**PRELIMINARY RESULTS** During the proof-of-concept phase, results were promising at both the national and store levels. The table below displays a summary of the store-level comparisons.

Descriptive Statistics for Differences between Store-Level Retailer 2012 NPD Data and 2012 Economic Census Data

| | Retailer 1 | Retailer 2 | Retailer 3 |
|---|---|---|---|
| Average difference between NPD and MRTS data | -0.08% | -0.64% | 0.56% |
| Average absolute difference between NPD data and MRTS data | 2.20% | 1.79% | 1.87% |
| Median difference between NPD and MRTS data | -0.05% | -0.55% | -0.01% |
| No. of months with absolute difference between NPD and MRTS data <1% | 32/48 | 35/48 | 24/48 |

*Disclaimer: Any views expressed in this paper are those of the authors and not necessarily those of the U.S. Census Bureau.*