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

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

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 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?

In 2017, a pilot project was undertaken by an Economic Directorate retail big data team along with the NPD Group, Inc., a privately held market information and business solutions company that captures point-of-sale transaction data from major retailers at the store and product level. In this pilot, we compared aggregated sales data from NPD at the product, store, and national levels to sales data reported by the retailer to our retail data products. The findings from this effort found that the NPD sales data aligned well at the national-level when compared to data reported to our Monthly Retail Trade Survey and the Annual Retail Trade Survey, and at the store-level when compared to 2012 Economic Census.

Building upon the success of the pilot, we are now expanding the effort to a production environment to include up to 100 more retailers with a focus on non-responding and high-burden retailers. This project has been identified as a pilot candidate for the new Census Bureau Data Lake which would allow data sharing across the enterprise.

This paper focuses on the potential gains and challenges that this expansion effort brings with it. With this data, we have the potential to reduce respondent burden, improve data quality, and create additional data products. In order to achieve those goals, we need to develop both an efficient and effective quality review process of the data and an automated, seamless transfer of the data to survey infrastructure all while operating under limited budgetary resources.