

FROM CHECKOUT TO DATA PRODUCT: USING RETAIL POINT-OF-SALE THIRD-PARTY DATA IN U.S. CENSUS BUREAU DATA PRODUCTS

FROM CHECKOUT

When NPD receives the feed from retailers each week, they cleanse the data. The retailer data are rich in store and product-level information and can be curated and aggregated to meet data users' business needs. For example, data users have the ability to know the total sales of product X in store location Y for retailer Z. Aggregated up across products, the user can know the total sales at store location Y and aggregated across stores, the total sales for retailer Z.



Every time a consumer purchases an item online or has a barcode scanned for purchase in a store, detailed data about that item is captured. We call this data "scanner data."

Scanner data is stored in aggregated feeds by SKU and these feeds can provide information about prices of the item, quantities of the item sold, and detailed product characteristics of the item. These aggregated data feeds do NOT contain any information about the individual transaction or the purchasers.

Sometimes retailers provide aggregated scanner data feeds to third-party businesses that conduct retail market research analysis. One of these companies is The NPD Group, Inc.

The Census Bureau recognized that data in the NPD retailer feeds may be similar to the data retailers are being asked to provide on Census Bureau survey instruments. These NPD data could be used to reduce respondent burden, help with survey nonresponse, and/or be used to create new data products.



To test the concept of using third-party data in place of data that is collected through traditional survey instruments, the Census Bureau first purchased store and product-level data for a small group of retailers from NPD and compared it data reported to Census Bureau surveys.

The NPD retailer data is run through a series of quality checks. First, national-level NPD retailer data is compared to the retailer data in the Monthly and Annual Retail Trade Surveys. Second, store-level NPD data is compared to the Economic Census to verify the number of store locations and product categories.

This comparison is straightforward when the retailer has reported to the survey, and less so when the retailer has not. When there is a large difference between the NPD data and Survey data, Census and NPD staff work together to validate the NPD data.



When the data track well, the data can be considered for use in Census Bureau data products.

Tabulated point-of-sale data at the national level for an individual retailer can be used to supplement existing surveys, including the Economic Census, the Monthly Retail Trade Survey and the Annual Retail Trade Survey, where the data can be used to help with nonresponse and validation.



These data can also be used to create new, blended data products like the Monthly State Retail sales where the point-of-sale data are used with existing survey and administrative data to create new estimates without any new data collections.

TO DATA PRODUCT

HOW THE DATA ARE USED

MITIGATE SURVEY NONRESPONSE

OVERVIEW: Point-of-sale retail data can capture retail sales data similar to what is collected on survey instruments. These data can mitigate sales data nonresponse in the Monthly Retail Trade Survey (MRTS), the Annual Retail Trade Survey (ARTS), and the Economic Census. These data are also useful for validating reported sales data for select retailers.

DATASETS: Monthly national-level and store-level datasets contain sales data by store location and product type for individual retailers who agree to allow NPD to share their data with the Census Bureau. The national level sales data are useful to the MRTS and the ARTS. Store-level and product-level sales are useful to the Economic Census. Retailers must agree to allow NPD to share their data with the Census Bureau.

NOTEABLE: Before using the store-level NPD data, we reviewed the data and assessed the quality using visual, statistical, and regression analysis between the NPD data and the data the retailer has reported to the Census Bureau either through the MRTS, the ARTS, or the Economic Census. In general, the NPD data often have a tight alignment to the data that retailers report to the monthly retail trade survey. Figure 1 shows a plot of the year-over-year percentage changes for both NPD data and Monthly Retail Trade Survey data for retailers with NPD data.

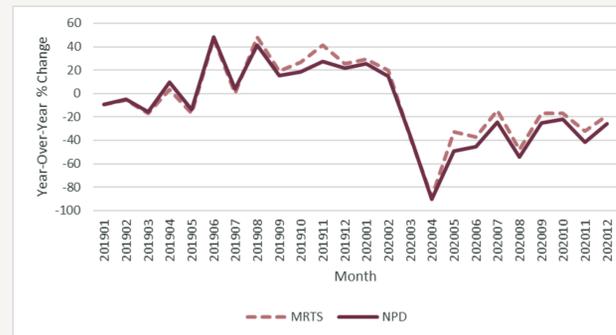


Figure 1: Year-Over-Year Percentage Changes Comparison of NPD Data and Corresponding MRTS Store-Level Data for Retailers with NPD data.

CREATE NEW DATA PRODUCTS

OVERVIEW: The Monthly State Retail Sales (MSRS) is a new experimental product that produces estimates using a composite estimator that blends third-party data with existing survey and administrative data. These data were requested by data users and were created with no new data collection. Previously, state-level retail sales data were only available on an annual basis every 5 year in the Economic Census. The use of third-party data allowed for the created of these monthly state-level data. Figure 2 shows a map of Total Retail Sales Excluding Nonstore Retailers for December 2020.

DATASETS: Monthly store-level NPD datasets contain sales data by store location, store numbers, and a postal code; these data are used in the imputation and estimation of the MSRS estimates. Monthly state-level datasets of aggregated sales for groupings of retailers specific by the Census Bureau; these data are used in the estimation of the MSRS data.

NOTEABLE: Store-level data can be difficult to obtain because they require retailer consent and are expensive. In order to obtain more data to improve the quality of the MSRS estimates, we worked with NPD to develop the state-level aggregate sales data for use in the MSRS data. This aggregation ensures the privacy of any one retailer's data but still captures valuable state-level sales information and does not require retailer consent to be included in the aggregation.

