

Overview of Monthly State Retail Sales Data

The Census Bureau has released the Monthly State Retail Sales (MSRS), an experimental data product featuring monthly state-level retail sales. These data are some of the most requested from our data users. Previously, state-level retail sales information was only available every five years as part of the Economic Census. The high-cost of survey collection and respondent burden has made survey-based state-level retail sales difficult and costly to obtain so these data are different from Census Bureau survey data. They are created using a composite model that incorporates Monthly Retail Trade Survey (MRTS) data, administrative data, and third-party data. This document provides an overview of this data product.

This is the first version of these experimental data; the Census Bureau plans to refine the series and invites users to provide feedback on how to improve this experimental product or on potential data sources to be used in the blended approach. Feedback can be shared by email at EID.Monthly.State.Retail@census.gov.

What are we publishing?

We are publishing year-over-year percent changes for each state and the District of Columbia for Total Retail Sales excluding Nonstore Retailers as well as for 11 three-digit retail subsectors as classified by the North American Industry Classification System (NAICS). These NAICS include Motor vehicle and parts dealers (NAICS 441), Furniture and Home Furnishing (NAICS 442), Electronics and Appliances (NAICS 443), Building Materials and Supplies Dealers (NAICS 444), Food and Beverage (NAICS 445), Health and Personal Care (NAICS 446), Gasoline Stations (NAICS 447), Clothing and Clothing Accessories (NAICS 448), Sporting Goods and Hobby (NAICS 451), General Merchandise (NAICS 452), and Miscellaneous Store Retailers (NAICS 453).

We are providing year-over-year percent changes back to January 2019. The state-level data is not adjusted for seasonal variation, trading-day differences, moving holidays or price changes.

Why are we not publishing Nonstore Retailers (NAICS 454)?

This publication generates geographic estimates based on the physical business location. With nonstore retailers, geographic performance is less tied to the physical location of the business than their store counterparts. Thus, we felt it would be misleading to try to publish geographic estimates for this industry. We will continue to work on a methodology with the hopes of releasing this data at a later date.

How are these state-level retail data produced?

The state-level retail sales estimates are composite estimates that combine synthetic estimates and hybrid estimates using third-party and directly-collected MRTS data as well as modeled establishment data. Figure 1 provides a simplified walkthrough of this methodology.

What are the limitations of the methodology?

The composite estimate is used because hybrid and synthetic estimates have different strengths and weaknesses. For the composite estimates, both the synthetic and hybrid estimates methods are based on the assumption that there is a strong relationship between retail sales and total annual payroll. Additionally, both estimators use MRTS data and therefore are subject to the same sampling and nonsampling errors associated with MRTS. Because MRTS is a sample-based survey designed for national estimates, geographic level MRTS data may not be representative of the true population at a state level.

Since the synthetic estimator is based primarily on total annual payroll and national MRTS estimates, any regional

or state seasonal patterns are not reflected in the estimates. It also does not directly use available establishment level data like the hybrid estimator does.

There are also limitations to the hybrid estimator. It assumes there is no measurement error from third-party or MRTS directly-collected data. Retailers with a single store location that are not in MRTS are not directly imputed but rather accounted for using a national industry level adjustment ratio. From all retailers with data in MRTS, only single location retailers or retailers with multiple locations within the same state are directly used in the calculation of the state-level composite estimates.

State sales estimates in each three-digit NAICS are calculated independently each month. Therefore, year-to-year percent changes may exhibit more variation if the coverage or percentage of composite estimator coming from the hybrid model changes between years.

What is the quality of the data?

These data are experimental and may not meet all of the quality standards of our official statistical products. To allow data users to assess the quality of the data, we are providing a variety of quality metrics including:

- **Standard Error:** Standard errors provide measures of variability for the year-to-year percent changes and can be used to construct confidence intervals when drawing inferences about the data.
- **Coverage:** This metric is produced for all monthly estimates, at the individual industry by state-level and at aggregated levels to show the proportion of the estimates that is directly collected either through MRTS or through a third-party data source. This metric also considers the proportion of the composite estimate coming from the hybrid estimate and will be lower the more the composite estimator has to rely on synthetic estimate, which does not use directly reported data. The quality of the model improves with better coverage.

These metrics are provided on the website at https://www.census.gov/retail/state_retail_sales.html.

When will these data be published?

The MSRS data use the final version of MRTS data as an input so the state-level data will be published approximately three months after month's end. We are exploring ways to publish this data sooner.

What improvements are planned for this product in the future?

- Publish sales levels in addition to year-over-year percent changes in the future.
- Develop a methodology to allocate nonstore sales to state-level geographies in an appropriate manner.
- Incorporate more store-location or state-level data that is directly collected either from a survey or third-party data to improve the quality of the data.

How can you share feedback on this data product?

Please share your feedback via email at EID.Monthly.State.Retail@census.gov.

For more information on Census Bureau experimental data products, please visit <https://www.census.gov/data/experimental-data-products.html>.

The data, quality metrics, and technical documentation for this release can be found at https://www.census.gov/retail/state_retail_sales.html.

The Census Bureau has reviewed this monthly data product for unauthorized disclosure of confidential information and has approved the disclosure avoidance practices applied. (Approval ID: CBDRB-FY23-004).

How are the state level retail sales data created?

Step 1 Calculate Synthetic Estimates

1

For each state and 3-digit NAICS combination, national Monthly Retail Trade Survey (MRTS) brick & mortar sales are allocated to states using administrative data (payroll) totals for each NAICS.

Estimate for each State & NAICS

$$= \text{National MRTS Sales for NAICS} * \frac{\text{Payroll for State \& NAICS}}{\text{National Payroll in NAICS}}$$

Step 2 Calculate Hybrid Estimates

2

For each state and 3-digit NAICS combination, we use available store-location or state-level data collected in MRTS or obtained through a third-party data provider to create state-level retail estimates.

Hybrid Estimates

	Data Included	Based on this Condition
=	MRTS brick & mortar sales	For MRTS retailers with more than one store location but all store locations are in the same state For MRTS retailers with only one store location
	+	
	Sum of third-party data brick & mortar sales	For MRTS retailers with more than one store location whose store-location or state-level data is available through a third-party
	+	
	Imputed brick & mortar sales	For MRTS retailers with more than one location who operate in multiple states and are not included in third-party data sales and information from third-party data
	+	
	Adjustment	To account for retailers with only one location that are not in MRTS or in the third-party data

Step 3 Calculate a weighted average of the two estimates

3

For each state and 3-digit NAICS combination, the final estimate is a weighted average and the weight is a ratio of the variances of the synthetic and hybrid estimates.

$$\text{Ratio} = \frac{\text{Variance (Synthetic Estimate)}}{\text{Variance(Synthetic Estimate) + Variance(Hybrid Estimate)}}$$

$$\text{Estimate} = (\text{Ratio} \times \text{Hybrid Estimate}) + ((1 - \text{Ratio}) \times \text{Synthetic Estimate})$$

The final state estimates are then adjusted to sum to the national MRTS total for each NAICS and year-over-year percent changes are calculated for publication.

Complete technical documentation available at: https://www.census.gov/retail/mrts/www/statedata/msrs_technical_documentation.pdf