

Investigating Alternative Data Sources to Reduce Respondent Burden in United States Census Bureau Retail Economic Data Products

Rebecca J. Hutchinson

Economic Directorate, United States Census Bureau

rebecca.j.hutchinson@census.gov

Disclaimer: Any views expressed are those of the author and not necessarily those of the United States Census Bureau.

Data user demands for more timely and more detailed data

Decline in respondent cooperation

Challenges facing the Economic Directorate

Changing economic landscape

Increasing costs of traditional survey data collection

Can alternative data sources help?

Point-of-sale data used to capture retail sales

Building permit data captured through API

Possible ways alternative data could be used

Capture new construction through satellite imagery

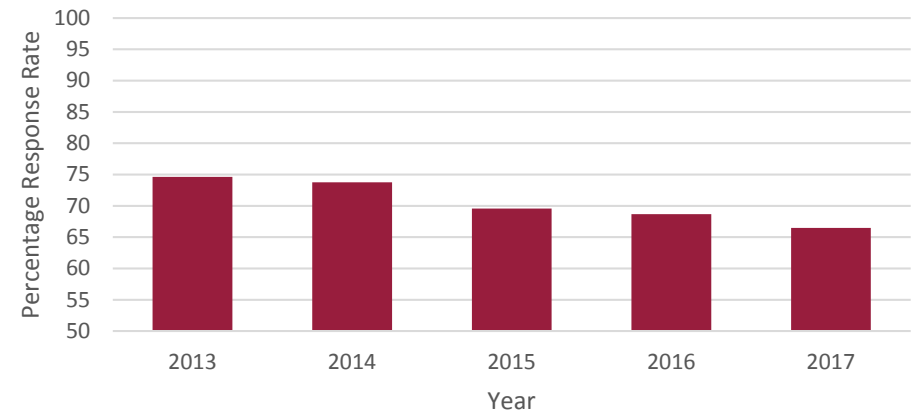
Web-scraping to capture publicly available financial filings



- Dynamic environment
- Innovative industry disruptors
- Evolution of online shopping

- Declining response rates

Average Monthly Response Rate -
Monthly Retail Trade Survey





Economic Census

Conducted every five years (years ending in '2' and '7')

Mandatory

Collects data at the establishment or store level on:

- Business characteristics
- Employment and payroll information
- Detail product-level sales information



Annual Retail Trade Survey

Conducted annually

Mandatory

Collects data at the company level

- Business characteristics
- E-commerce
- Sales & Inventories
- Expenses



**How are
Retail Sales
currently
measured?**

Conducted monthly

Voluntary

Collects data at the company level on:

- Limited business characteristics
- Sales
- Inventories
- E-commerce



Monthly Retail Trade Survey

Conducted monthly

Voluntary

Collects data at the company level on:

- Limited business characteristics
- Sales
- E-commerce



Advanced Monthly Retail Trade Survey

What data items do we need?

If point-of-sale data captures every sale made in store or online, would the sum of all these sales equal total retail sales for a given retailer?

Do we get the data from a third-party vendor?

What IT resources do we need?

How do we determine the quality of the data?

How do we implement without adding to analyst workload?

Do we get the data directly from retailers?

Selected through the official government acquisitions process, the NPD Group, Inc. (NPD) was chosen as a third-party data source vendor.



- Captures point-of-sale data from over 1,300 retailers representing 300,000 stores and e-commerce platforms worldwide.
- Processes data for many product categories including apparel, small appliances, automotive, beauty, fashion accessories, consumer electronics, footwear, office supplies, toys, video games, and jewelry and watches.
- Creates an unclassified buckets for categories that it does not process.
- No information about individual purchasers or transactions is collected.
- Coupon values, discounts, sales tax, and shipping & handling are excluded from the sales.



- How well do national-level sales data tabulated from the point-of-sale data compare to data being tabbed for retailers in monthly retail surveys?
- If the data aligned well for retailers who reported, how is the quality of the point-of-sale data for those retailers who do not report to survey determined?
- How well do store-level sales and location data tabulated from the point-of-sale data compare to data that retailers reported the 2012 Economic Census and 2017 Economic Census?
- How well do the product categories in the point-of-sale data align to the North American Product Classification System used in the 2017 Economic Census?
- If the mapping is possible, how well do the product sales compare between the NPD data and Economic Census data?



- Monthly Retail Trade Survey
- Annual Retail Trade Survey
- Economic Census Data
- Public Financial Filings
- Administrative Data



- Economic Census



- Economic Census



2017

- Proof-of-concept effort with data from 3 retailers who were good and consistent survey reporters.
- Expanded to include 13 more retailers including non-reporters

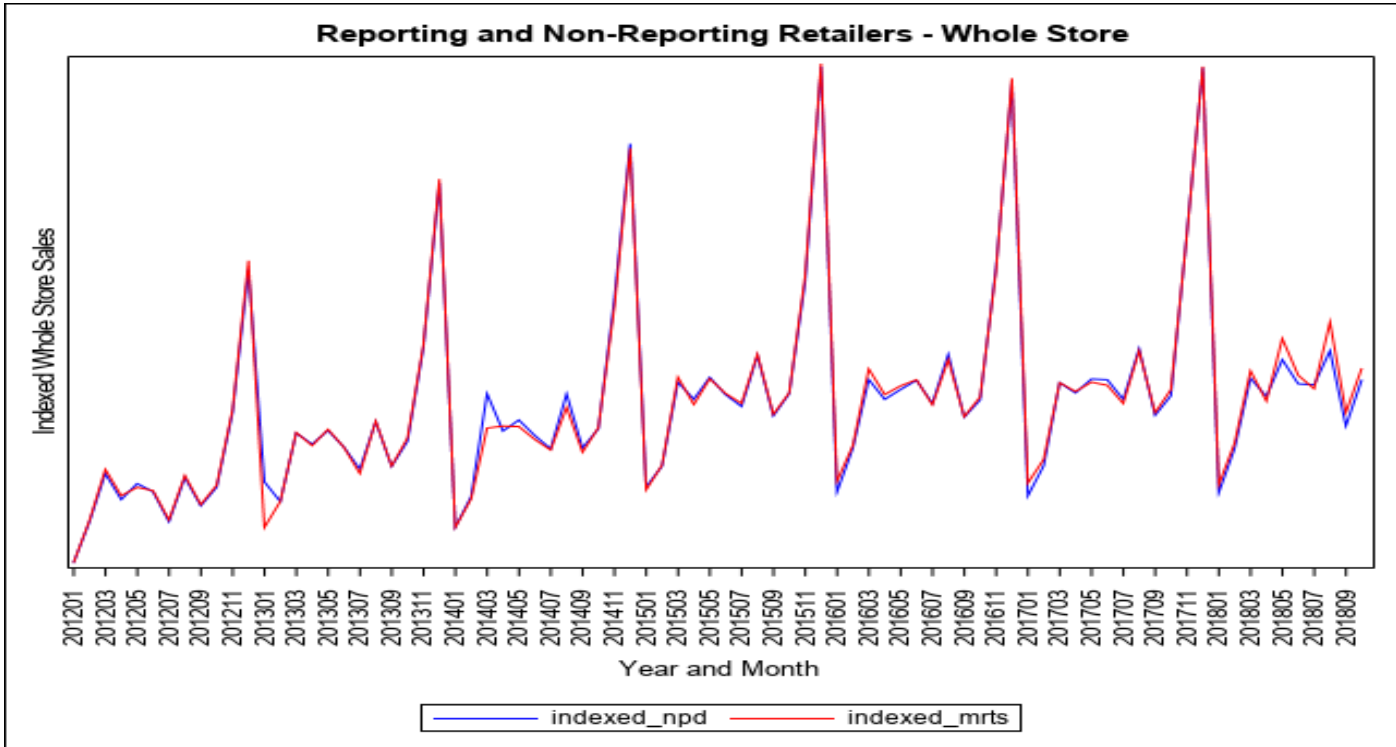
2018

- Contract awarded for purchase data for 60 retailers in FY 2019

2019

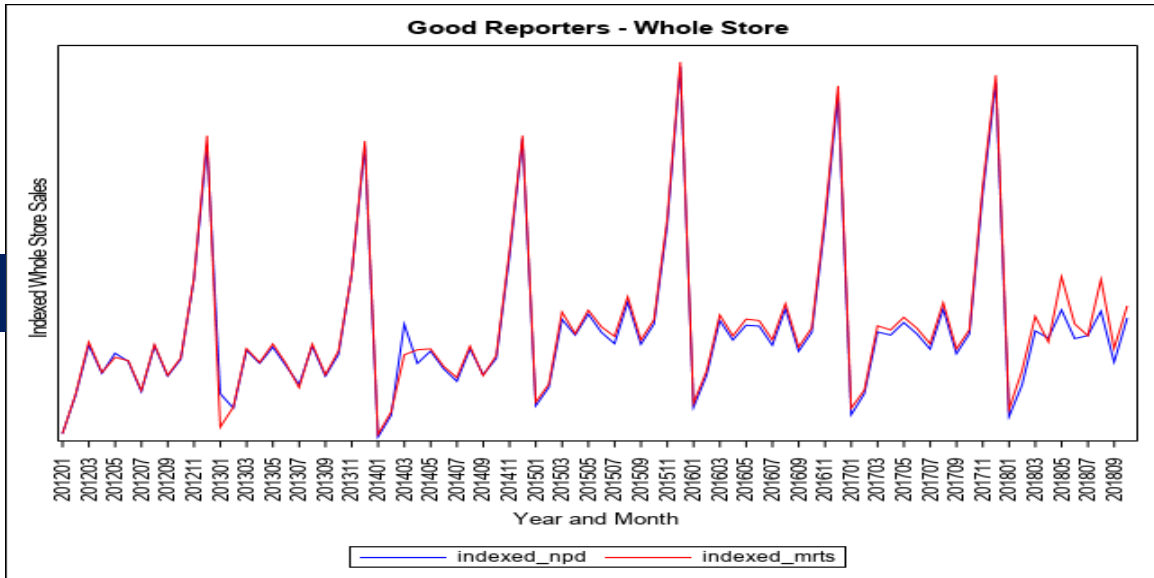
- To date, 20 retailers have agreed to share data through NPD.

How well do national-level sales data tabulated from the point-of-sale data compare to tabulated for the retailers in the Monthly Retail Trade Survey?

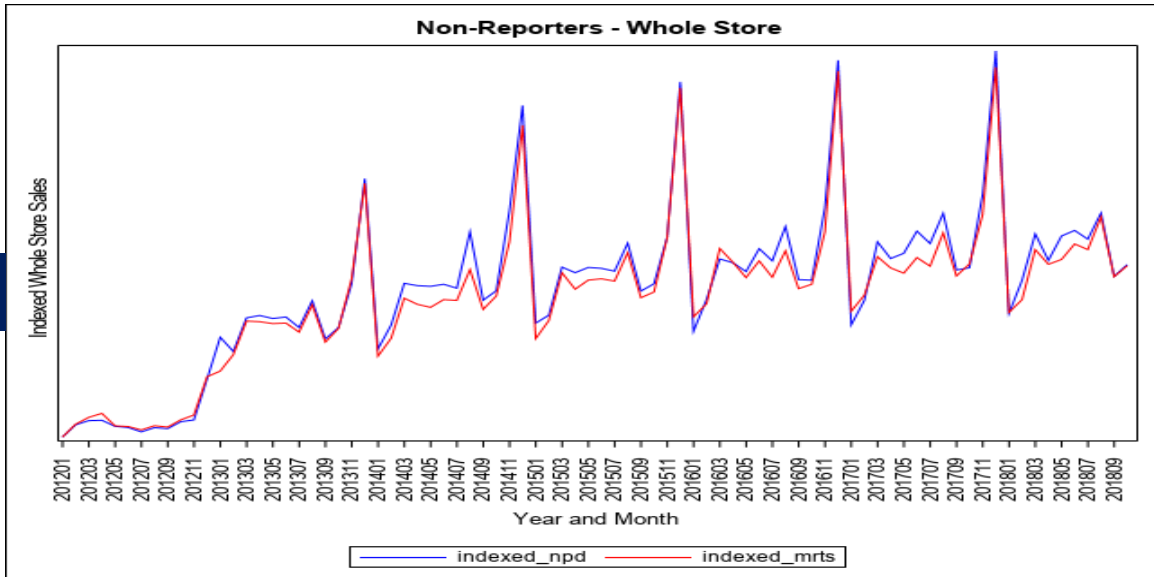




Good Reporters



Non-reporters





NPD delivered in time for the Advance Monthly Retail Trade Survey estimates.

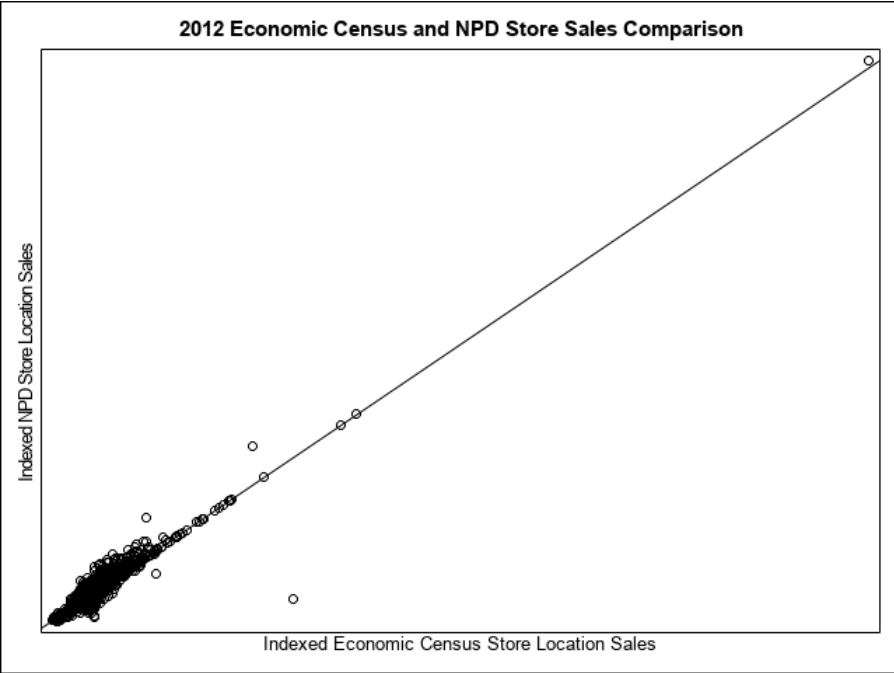
- Using NPD data for some retailers that do NOT report in the MRTS and ARTS estimates.
- Using NPD data for retailers that do report to verify reported data.



- **Identifying sources of data discrepancies**
- **Communication of questions**
- **Improving data ingest process**
- **Survey staff buy-in to the effort**

How well do store-level sales and location data tabulated from the point-of-sale data compare to data that retailers reported the 2012 Economic Census?

- Store-level data has the potential to relieve tremendous reporting burden on the Economic Census.
- Store number variable was critical to successful matching.
- Store location match rate between NPD and 2012 Economic Census is 99%.

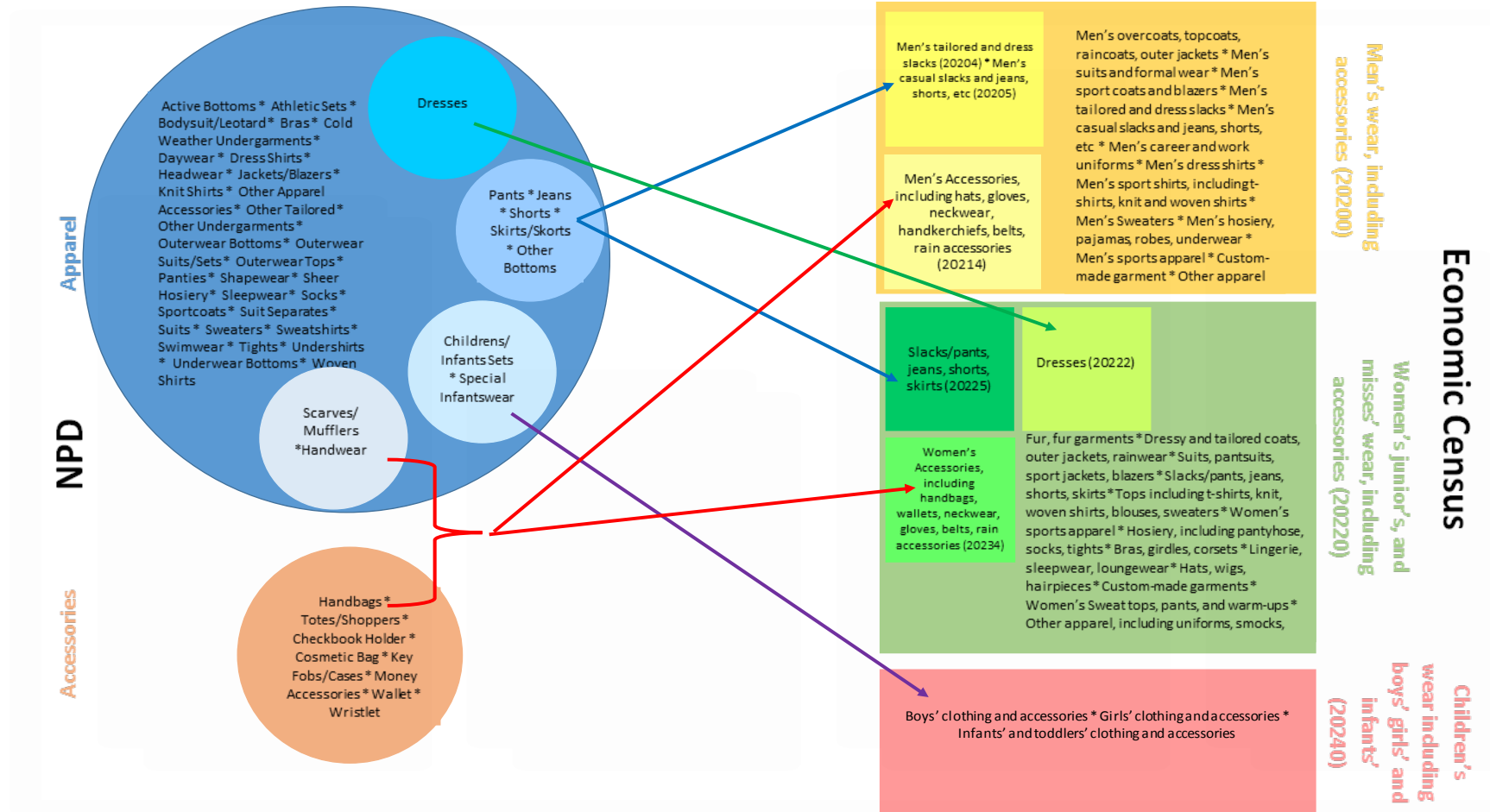




Because of their different purposes and data uses, the NPD and Census Bureau's product categories are different.

- Point-of-sale data from NPD is collected at the stock-keeping unit level (SKU). SKUs are then assigned detailed product attributes placed into broader categories including apparel, small appliances, automotive, beauty, fashion accessories, consumer electronics, footwear, office supplies, toys, video games, and jewelry.
- Through the 2012 Economic Census, the Census Bureau used its own set of broad and detailed industry product categories. Beginning with the 2017 Economic Census, the North American Product Classification System (NAPCS) was implemented.

How well do the product categories in the point-of-sale data align to the product categories used in the 2012 Economic Census?



During the summer of 2018, a mapping of the full NPD product catalog to NAPCS was completed in cooperation with NPD and Census Bureau Classification staff.



NPD Product Catalog					NAPCS Classification	
Broad Category	Super Category	Category	Class	Sub Class	NAPCS Code	Description
Apparel	Total Bottoms	Total Pants	Men's Pants	Men's Casual Pants	143010101	Retailing services for men's clothing
Apparel	Total Bottoms	Total Pants	Men's Pants	Men's Dress Pants		
Apparel	Total Bottoms	Total Pants	Women's Pants	Women's Casual Pants	143010102	Retailing services for women's clothing
Apparel	Total Bottoms	Total Pants	Women's Pants	Women's Dress Pants		
Apparel	Total Bottoms	Total Pants	Women's Pants	Women's Leggings		
Apparel	Total Bottoms	Total Pants	Boy's Pants	Boy's Casual Pants	143010103	Retailing services for children's wear, including boys', girls', and infants' and toddlers' clothing and accessories retailing services
Apparel	Total Bottoms	Total Pants	Boy's Pants	Boy's Dress Pants		
Apparel	Total Bottoms	Total Pants	Girl's Pants	Girl's Casual Pants		
Apparel	Total Bottoms	Total Pants	Girl's Pants	Girl's Dress Pants		
Apparel	Total Bottoms	Total Pants	Girl's Pants	Girl's Leggings		
Apparel	Total Bottoms	Total Pants	Infant/Toddler Pants	Male Infant/Toddler Casual Pants		
Apparel	Total Bottoms	Total Pants	Infant/Toddler Pants	Female Infant/Toddler Casual Pants		
Apparel	Total Bottoms	Total Pants	Infant/Toddler Pants	Male Infant/Toddler Dress Pants		
Apparel	Total Bottoms	Total Pants	Infant/Toddler Pants	Female Infant/Toddler Dress Pants		
Apparel	Total Bottoms	Total Pants	Infant/Toddler Pants	Male Infant/Toddler Leggings		
Apparel	Total Bottoms	Total Pants	Infant/Toddler Pants	Female Infant/Toddler Leggings		



Men's Clothing	Sales Data Reported to 2017 Economic Census	Sales Data Available in NPD feeds
Outerwear coats, jackets, windbreakers, and similar		X
Suits and formal wear		
Sport coats and blazers		
Tailored and dress slacks		
Casual slacks and jeans, walking shorts, etc.		X
Dress shirts		X
Sports shirts, including t-shirts, knit and woven shirts, etc.		X
Sweaters		X
Sweat tops, pants, and warm-ups	X	X
Underwear, nightwear, and hosiery		X
Career and work uniforms		X
Sports apparel, including tennis, golf, jogging, swimming, skiing, camping, fishing, hiking, and other rugged outer and exercise apparel		X
Other men's wear		



- Obtaining additional information on the Unclassified bucket in the NPD data
- Developed a standard format output to allow tabulated NPD product-level data to be loaded to Economic Census database.
- Developing other data products from the product-level data



Challenges

- **Cost**
- **Availability of other data items**
- **Modifying collection efforts**

