Reducing Survey Burden Through Third-Party Data Sources

Rebecca J. Hutchinson
Economic Directorate, United States Census Bureau
UNECE Workshop on Survey Collection

Disclaimer: Any views expressed in this presentation are those of the author and not necessarily those of the U.S. Census Bureau.
Focus on using alternative sources of data

In January 2015, an Economic Directorate Retail Big Data Team was tasked with researching how can we leverage alternative data sources in conjunction with existing monthly and annual retail trade surveys and the Economic Census to:

- Offer greater insight into the nation’s economy through detailed geographic and industry-level estimates
- Reduce respondent burden
- Provide more timely data products
- Improve efficiency and quality throughout the survey life cycle
- Enhance current methodologies
Risks and challenges with third-party data

- Cost
- Transparency
- Confidentiality
- Consistency
- General data quality
Proof-of-concept effort: Respondent burden

Evolving retail economy generates great interest in official retail statistics

Declining response rates due to respondent burden

Can third-party data help reduce respondent burden while still maintaining the high quality of official statistics?
Overview of Retail Data Products

**Economic Census**
- Conducted every five years (years ending in ‘2’ and ‘7’)
- Mandatory
- Collects data at the establishment or store level
- Data collected include:
  - Business characteristics
  - Employment and payroll information
  - Detail product-level sales information

**Annual Retail Trade Survey (ARTS)**
- Conducted annually
- Mandatory
- Sampled from frame created by the Economic Census
- Collects data at the company level
- Data collected include:
  - Business characteristics
  - E-commerce
  - Sales
  - Inventories
  - Expenses

**Monthly Retail Trade Survey (MRTS)**
- Conducted monthly
- Voluntary
- Subsampled from the Annual Retail Trade Survey
- Collects data at the company level
- Data collected include:
  - Limited business characteristics
  - Reporting period information
  - Sales
  - E-commerce
Example of Monthly Retail Respondent Burden

2015 Form Burden for Multi-Unit MRTS Retailers

- Most MRTS multi-units receive no more than five forms
- Does not capture any forms retailers receive from other government agencies
Identifying a third-party data source

The NPD Group, Inc. obtains aggregated point-of-sale data feeds from over 1,250 retailers representing 300,000 stores and e-commerce.

Industries covered include but not limited to:

- Apparel
- Automotive
- Beauty
- Books
- Consumer electronics
- Diamonds
- Fashion accessories
- Footwear
- Jewelry/watches
- Office supplies
- Small appliances
- Toys
- Office supplies
What is NPD collecting?

Data feeds include:

- Product identifier
- Number of units sold
- Product sales in dollars
- Average price sold
- Total store sales in dollars
- Week ending date
What isn’t NPD collecting?

- No information about individual transactions or purchasers is collected
- Price reduction and coupon values are filtered out before feeds are created
- Sales tax and shipping and handling are excluded
Potential uses for the NPD data

The NPD data sets have the potential to:

- Reduce respondent burden in the Monthly and Annual surveys and in the Economic Census
- Provide more timely product-level estimates to supplement the Economic Census
Proof-of-concept project

Purchased monthly data sets for 3 retailers that are consistent, good reporters to our monthly and annual retail surveys and the Economic Census.

- Compare NPD data at the national and store levels to survey and Census data
- Compare NPD product lines to Econ Census product lines
Structure of data sets

Each data set includes monthly observations from 2012-2015 including:

- Store number
- Store zip code
- Brick & Mortar/E-commerce identifier
- Product category and subcategory
- Sales by product subcategory
National total stores sales line up well

![Whole Store Indexed Sales Comparison for Selected Retailers](chart)

January 2012=1.0

Indexed Consumer Spend

NPD  MRTS
Successful matching of store level data

<table>
<thead>
<tr>
<th></th>
<th>Retailer 1</th>
<th>Retailer 2</th>
<th>Retailer 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>NPD/2012 Econ Census Store Match Rate</td>
<td>100%</td>
<td>99.9%</td>
<td>100%</td>
</tr>
<tr>
<td>Average Percentage Difference in Store Sales Between NPD &amp; 2012 Econ Census</td>
<td>-2.36</td>
<td>-1.22</td>
<td>0.96</td>
</tr>
<tr>
<td>Median Percentage Difference in Store Sales Between NPD &amp; 2012 Econ Census</td>
<td>-2.18</td>
<td>-1.18</td>
<td>0.31</td>
</tr>
</tbody>
</table>
Product line matching is complex

NPD

NPD’s business model requires classifying products at a SKU level basis.

If the product falls into a category they don’t market, NPD places dollar amount in an unclassified bucket.

Economic Census

Product level data is captured in the 2012 Economic Census using the Census specific product line codes.
Apparel product line match-up

United States Census Bureau
Economics and Statistics Administration
U.S. CENSUS BUREAU
census.gov
What’s next for this work?

Part 1: Proof-of-Concept
Obtain data sets for 3 companies that are consistent, good reporters to our monthly and annual retail surveys and the Economic Census
- Compare NPD data at the national and store levels to survey and Census data
- Determine feasibility of mapping NPD product lines to Econ Census product lines

Part 2: Survey Response
Obtain data sets for companies that are
- non-respondents to our monthly and annual retail surveys
- included in multiple surveys and have a high respondent burden
- Determine the impact on estimates of using alternative data source to reduce burden.
- Conduct a validation exercise on current MRTS imputation methodology

Part 3: Econ Census
Obtain 2017 data sets for companies within a single NAICS that has good NPD coverage
- Work with NPD and Census Classification staff to map NPD products lines to Econ Census product lines
- Generate Econ Census store-level and product-level estimates and compare to collected Econ Census data when available
Please reach out with any questions or comments.

- Rebecca Hutchinson
  
  rebecca.j.hutchinson@census.gov