

Federal Economic Statistics Advisory Committee

# Economic Directorate's Retail Big Data Overview

June 10, 2016



U.S. Department of Commerce  
Economics and Statistics Administration  
U.S. CENSUS BUREAU  
census.gov

DRAFT – FOR INTERNAL USE ONLY



# Economic Directorate Retail Big Data Projects

## Goal

Leverage Big Data sources in conjunction with existing survey data to

- Provide more timely data products
- Offer greater insight into the nation's economy through detailed geographic and industry-level estimates
- Improve efficiency and quality of processing throughout the survey life cycle

# Economic Directorate Retail Big Data Projects

## Currently Active Projects

- Use of third-party data to add detail to Retail Trade survey data
  - NPD
  - Palantir/First Data
- Passive Data Collection

# Economic Directorate Retail Big Data Projects

## NPD

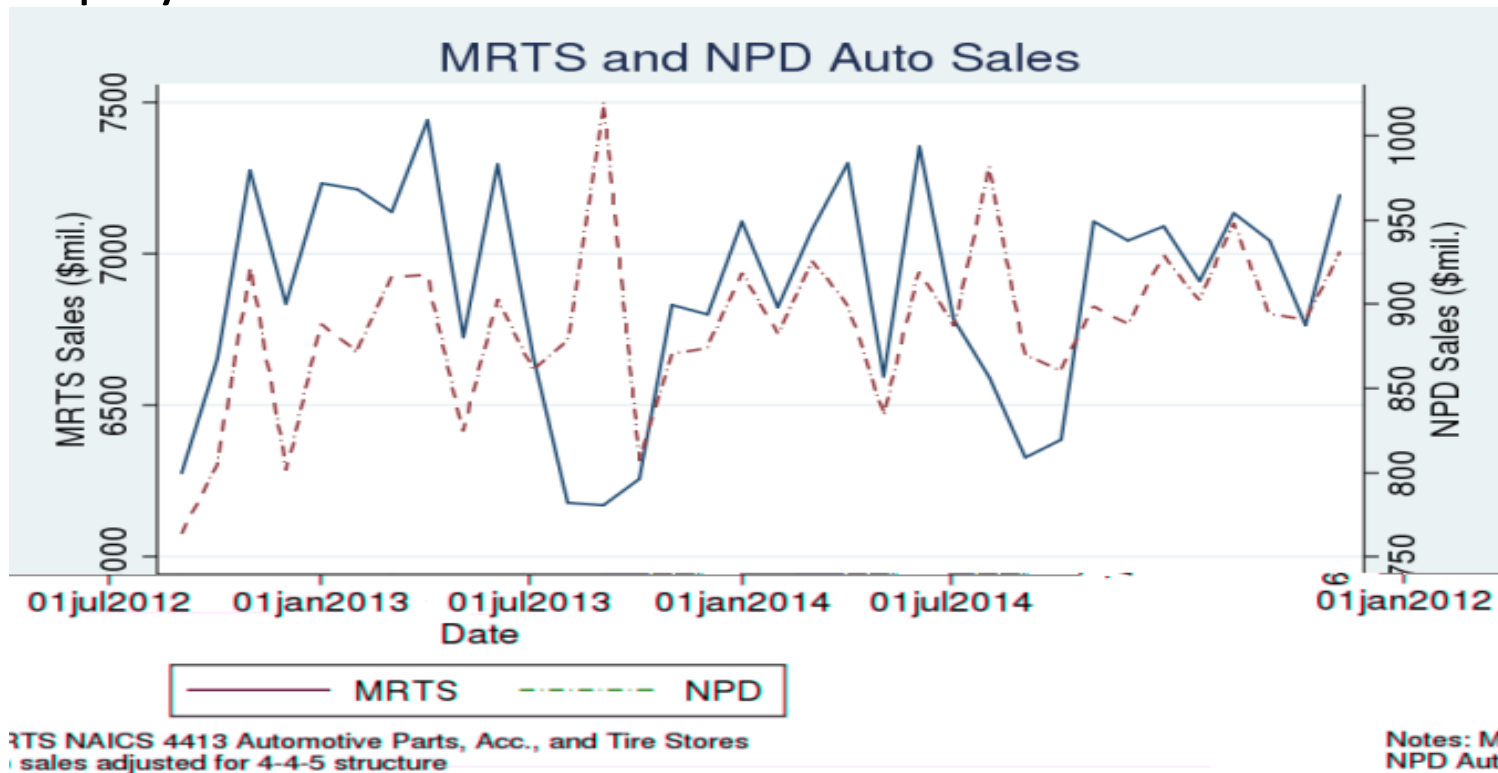
### NPD

- Purchased a dataset of point-of-sales transactions from January 2012 through December 2014
- Explored two datasets
  - Auto Parts
  - Jewelry and Watch data
- Obtained geographic detail at the Nielsen Designated Market Area (DMA) level

# Economic Directorate Retail Big Data Projects

## NPD

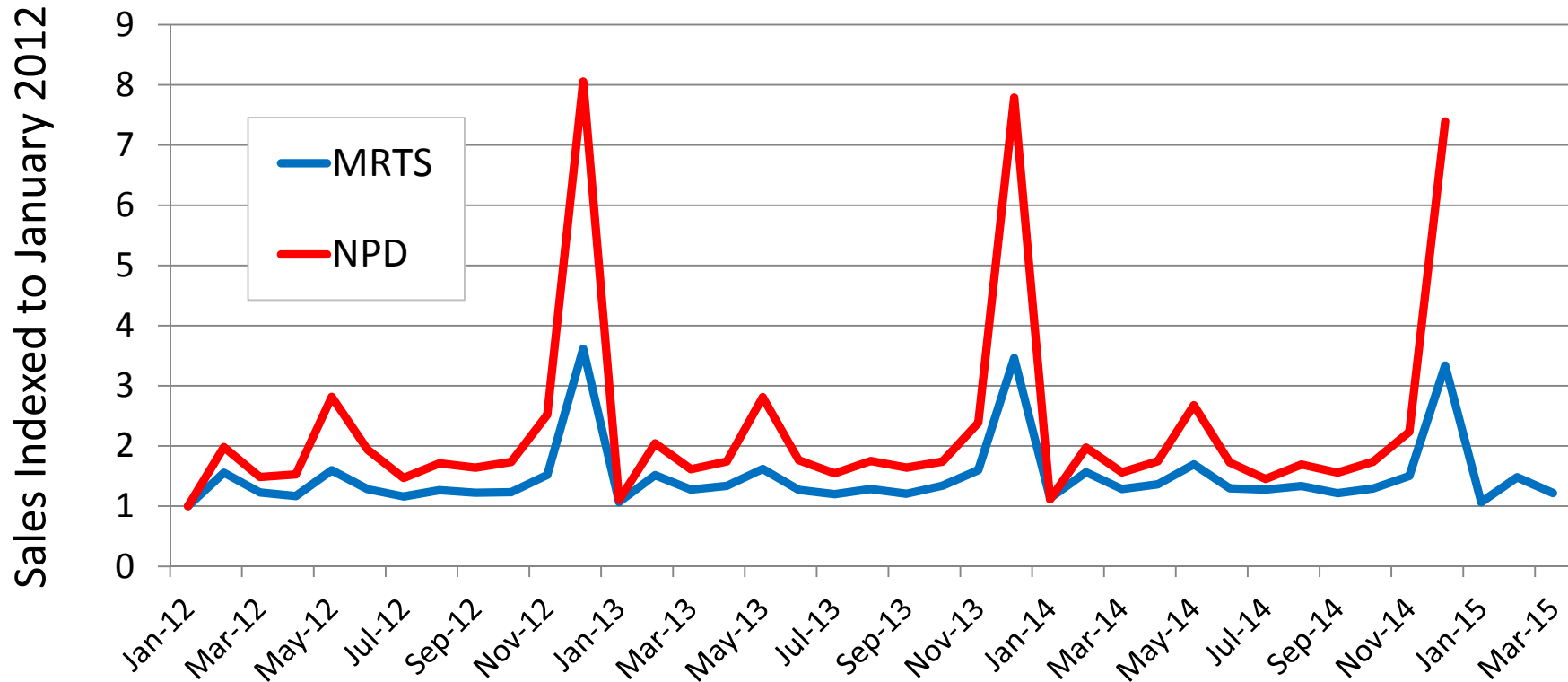
NPD Auto Parts and Monthly Retail Trade Survey did not display similar trends.



# Economic Directorate Big Data Projects

## NPD

NPD Jewelry & Watches and Monthly Retail Trade Survey did display similar trends, but not levels.



# Economic Directorate Big Data Projects

## NPD

### Recommendations for the use of NPD data:

- For geographic detail, Census needs data based on zip code, not Designated Market Area.
- Data sets need to be standardized to the same level of geography and detail.
- Product line data that align with Census Bureau product lines would be more useful than what was obtained

# Economic Directorate Big Data Projects

## Palantir/First Data

Collaborative short-term exploratory project involving Palantir, First Data, Bureau of Economic Analysis, and Census Bureau

### First Data's consumer spending data

- Cover 58 billion transactions annually
- Capture about 45% of all point-of-sale transactions
- Capture credit, debit, and prepaid gift card transactions but not cash
- Cover five states for this pilot project

### Palantir's software tool

- Updated daily with retail transactions
- Custom dashboards
- Environment for using R and Python

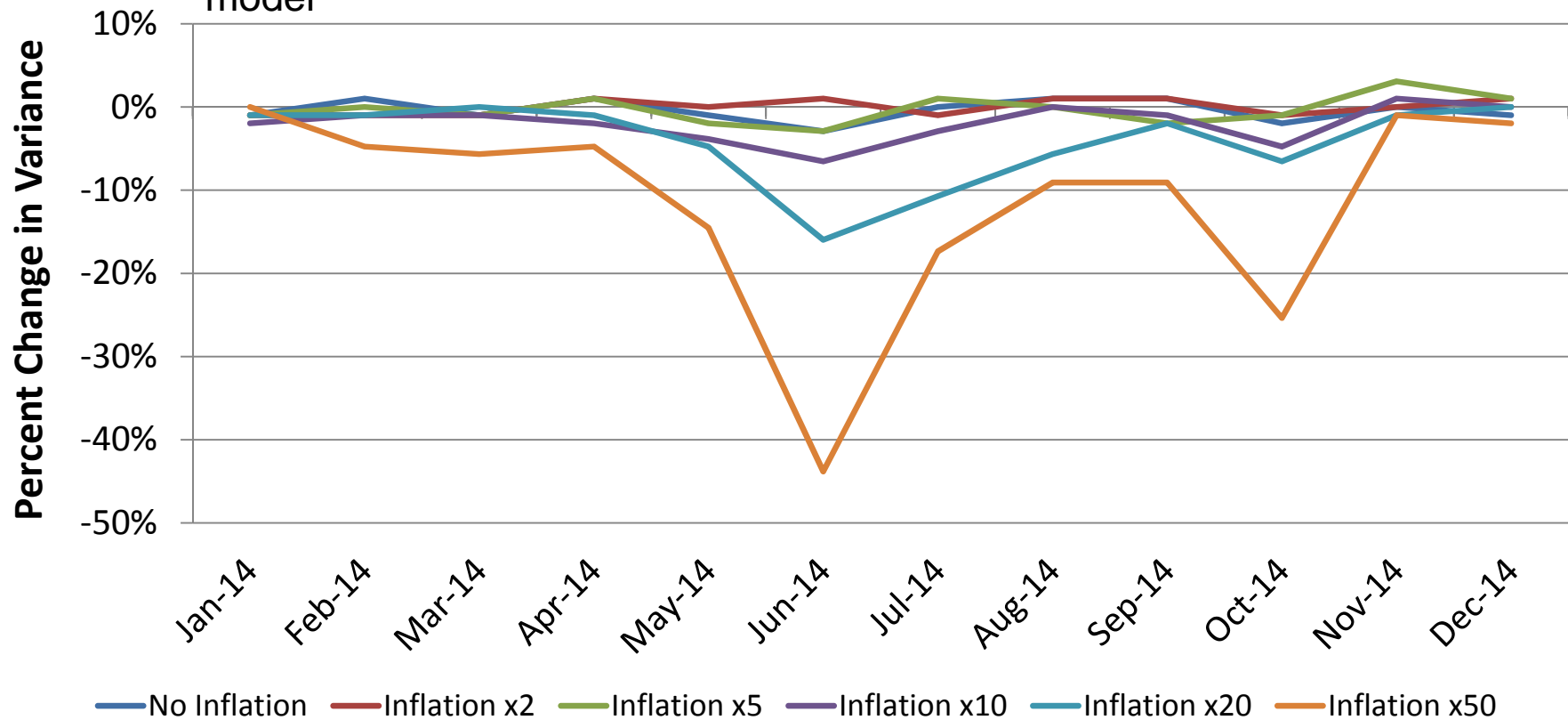


# Economic Directorate Retail Big Data Projects

## Palantir/First Data

### Building Small Area Estimation Models

Percent change in model variance over original variance for national-level model



# Economic Directorate Retail Big Data Projects

## Palantir/First Data

### Examining trading day weight calculations and holiday adjustments

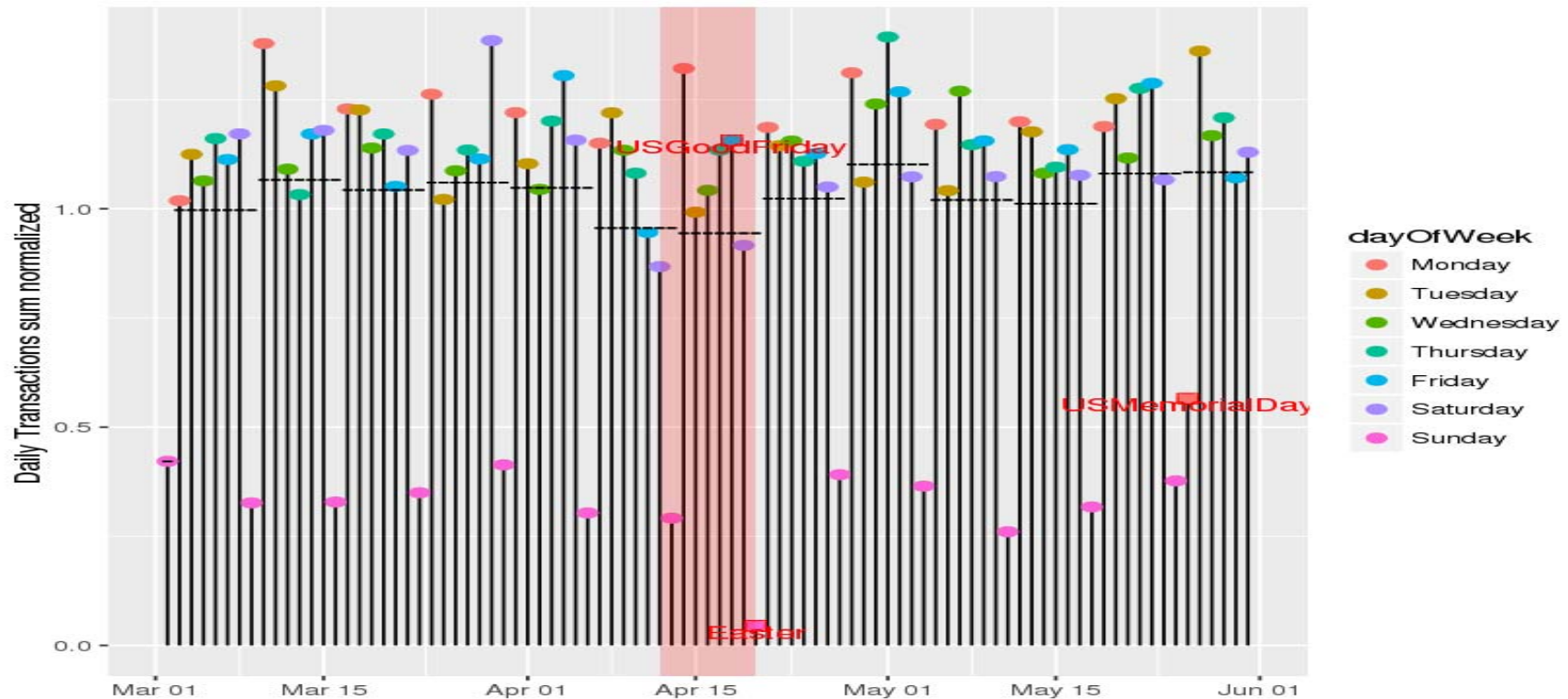
- Using daily data seasonal models developed by the US Census Bureau's Tucker McElroy and Brian Monsell
- Comparing daily data modeled from credit card output to current X-13 generated trading day weights and looking for areas of improvement
- Modeling holiday effects for Super Bowl Sunday, Chinese New Year, Easter Sunday, Ramadan, Labor Day, and Cyber Monday

# Economic Directorate Retail Big Data Projects

## Palantir/First Data

Daily data has revealed an Easter Sunday holiday effect

Appliance, Television, and Other Electronics Store  
(NAICS Code 44311) - 2014



# Economic Directorate Retail Big Data Projects

## Passive Data Collection

- Goal
  - Searching for opportunities to receive company data feeds that we can use across all of our surveys
- Two approaches
  - Reaching out to companies directly
  - Possibly partnering with a third-party source to determine if their respondents will agree to share their data with Census

# Economic Directorate Big Data Projects

## Next Steps

- More Timely/Granular Data Products
  - Continue to research third party data sources
  - Complete our work with Palantir/First Data and write the summary document
  - Review submissions to our Request for Information (RFI) to identify additional possibilities
- Passive Data Collection
  - Continue searching for opportunities to receive data feeds directly from large companies that we can use across all of our surveys
  - Research the quality of 3<sup>rd</sup> party sources for company level data

# Questions for the Committee

- In building our Fay-Herriot models, we are challenged with not having direct Monthly Retail Trade Survey estimates at a state-by-industry level. We are considering alternative inputs with known limitations; the variances associated with the alternative inputs are too small and the models are not offering as much benefit as they should. Does the Committee have any suggestions?
- Does the Committee have any recommendations on additional data sources for us to examine, particularly for our Retail Trade research?
- Does the Committee have any recommendations on the direction that we should follow in looking to reduce respondent burden – passive collection or other suggestions?
- Are there concerns with how we should blend official statistics and third-party data?