

Exploring New Sources of Spending Data: First Data Transactions Data

Adi Aladangady (Federal Reserve Board)
Dan Moulton (Palantir)
December 9, 2016



GOALS FOR USING TRANSACTION DATA AT THE FED

1. Produce a **timely, independent measure** of spending to improve our analysis and forecasting of national economic activity with a focus on Retail Sales
2. Use detailed cuts of the data to **improve the granularity of our understanding** of consumer spending behavior at the state and national level
3. Access and analyze anonymized aggregates of a new data set in a way that is **not a black box**

LESSONS LEARNED FROM PRIOR WORK WITH TRANSACTION DATA

- Data cleaning, filtering, and seasonal adjustment are time consuming
- Requesting data from private companies was an iterative process
- Variation in the ability/willingness of different data vendors to work toward a data product that met our needs
- Data received needs to ensure anonymity of individual observations and sensitive business information; understanding transformations applied to the raw data is important

FIRST DATA/PALANTIR: WHAT IS IT?

- First Data – An electronic payments processing company that provides merchants with the technology to process credit, debit, and electronic payment transactions
- Raw data is daily frequency, swipe-level transactions by merchant
 - Merchants are classified by MCC
 - Available to Board 3-5 days after transaction
 - Covers roughly 45% of national “core” retail sales
- Palantir constructs transforms to anonymize aggregates of transactions
 - Transparency – Underlying code for transforms available to Fed
 - Collaboration – Weekly engagement model to iterate on methodology

PROVIDING TRANSPARENCY INTO DATA PROVENANCE

Monocle

Layout Tools Expand Select Undo/Redo Find More

Dataset Details

Summary Columns Rows Code Pipemon

```
1 package pipeline.transform.first_data.national_series.chained;
2 import common.DateParsing;
3
4 newTable ("yoy_naics_national_transform") {
5   startWith "pipeline/cleaned/first_data/national_series/yoy_naics_national_cleaned"
6   choose columns: ["month", "yoy_estimate", "mom_estimate", "naics_4", "chain_method"]
7   updateColumn "month", function: DateParsing.parseDate, args: ["month"]
8   updateColumn "publication_date", function: DateParsing.parseDate, args: ["publica
9
10  def double_columns = ["yoy_estimate", "mom_estimate"]
11  double_columns.each{(column -> updateColumn column,
12    function: { v ->
13      if (v==null)
14        return null
15      return Double.parseDouble(v)
16    }, args: [column, type: "Double"]})
17
18
19 }
```

Monocle

Layout Tools Expand Select Undo/Redo Find More

Legend Color Node Color Options

- Raw Dataset (35)
- Mesa Dataset (150)
- Contour Dataset (10)
- Foundry Export Job (4)

199 datasets selected

Copy Names

Histogram

BRANCH	# DATASETS
master	199

DATASET SIZE	# DATASETS
Less than 100MB	167
100MB to 1GB	23
1GB to 10GB	9

DATASET TYPE	# DATASETS
Contour Dataset	10
Foundry Export Job	4
Mesa Dataset	150
Raw Dataset	35

FILE COUNT	# DATASETS
Less than 10	119
10 to 99	68
100 to 999	5
1,000 to 9,999	7

FREQUENT COLUMNS	# DATASETS
geo	56
naics4	54
txn_sum	53
txn_count	49
merchant_count	45

Show More (5/292 rows)

FREQUENT FOLDER PATHS	# DATASETS
pipeline/analytic/first_data/slate_tables	15
pipeline/cleaned/first_data/trucker_index	9
pipeline/analytic/first_data/trucker_index	9
pipeline/transform/first_data/trucker_index	9
pipeline/raw/first_data/trucker_index	9

LAST BUILT	# DATASETS
Last 6 hours	3

Last synced with Foundry at 2:23pm 12/05/2016



DATA + ANALYSIS IN ONE PLACE

All files > Users > Shifrah.R.Aron-Dine@frb.gov > Weather > Hurricane_Mathew - Kernel Type: ir

```
1 ##### READ ME #####  
2 # Lag 0 corresponds to 10/06/2016 #  
3 # possible inputs for NAICS: #  
4 # 442,443,445,446,447,448,451,452,453,"RSG" #  
5 # possible inputs for LAG: #  
6 # -4,-3,-2,-1,0,1,2 #  
7 #####
```

Run `impulse_response_by_day` with arguments:

LAG:

NAICS:

HURRICANE MATHEW IMPULSE RESPONSE RETAIL SALES GROUP

Percent Deviation in Sales

100
50
0
-50
-100



CONSTRUCTING AGGREGATE INDEXES

In its raw form, FD's transaction data measures the activity of its business, not the U.S. economy.

Working with >140 billion rows, we had to define how to isolate 'good merchant behavior' to build aggregates only from merchants representing real economic activity. This work included:

- Controlling for **merchant births and death** with 'chaining' methodology to reflect real entry/exit of sample growth; additionally, First Data acquisitions of new platforms results in large spikes
- Establishing a **merchant-to-industry mapping** from MCCs to 4-digit NAICS; this is particularly hard for multi-NAICS firms and e-commerce
- Applying **seasonal adjustment** to a short time series (2012-present) where X-13 type filters have limited use; we currently match NSA 12-month growth and use Census SA factors

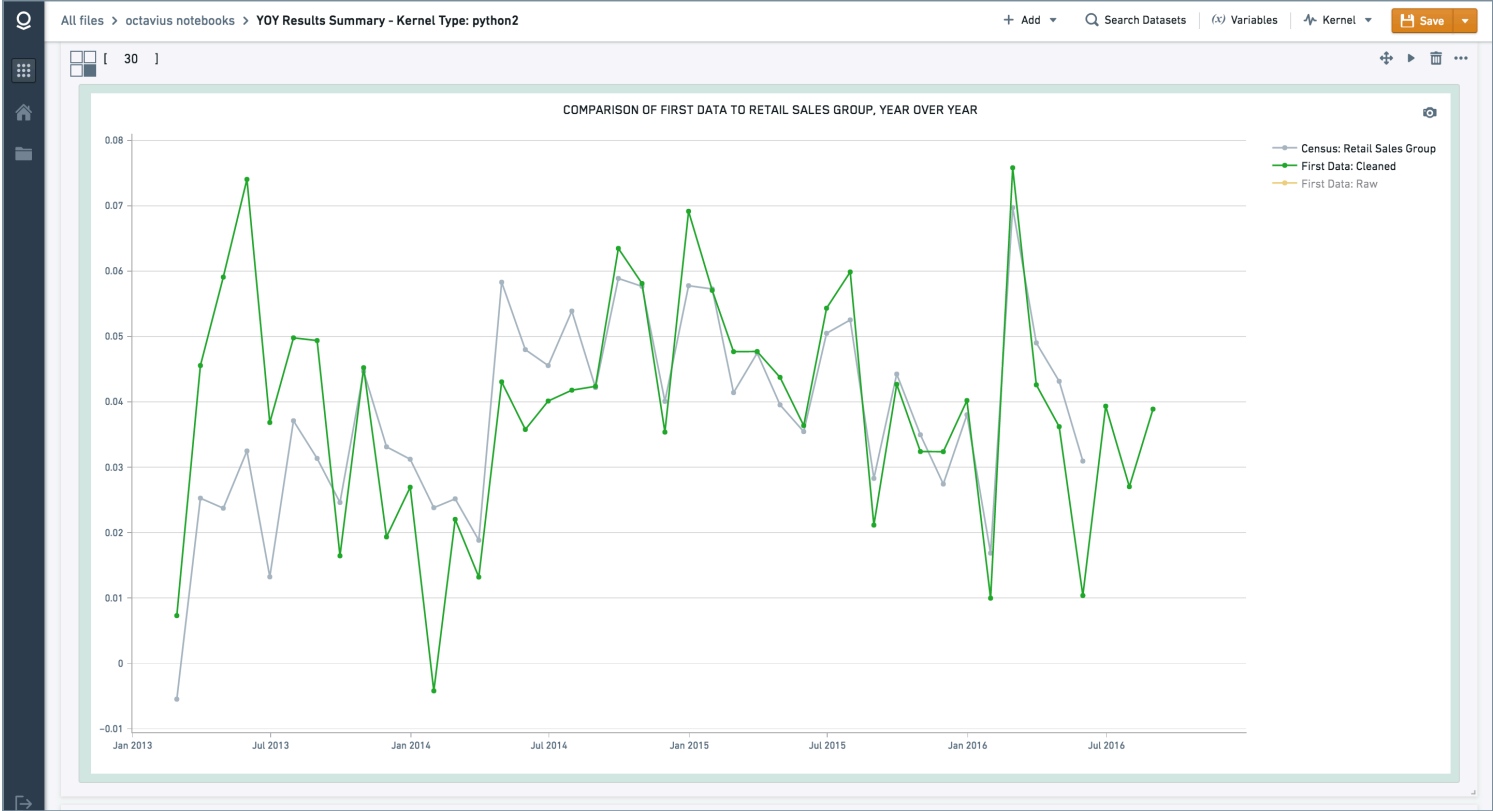
FIRST DATA/PALANTIR: HOW WELL DID IT WORK?

Raw Data vs Census Retail Sales (12mo change, NSA)



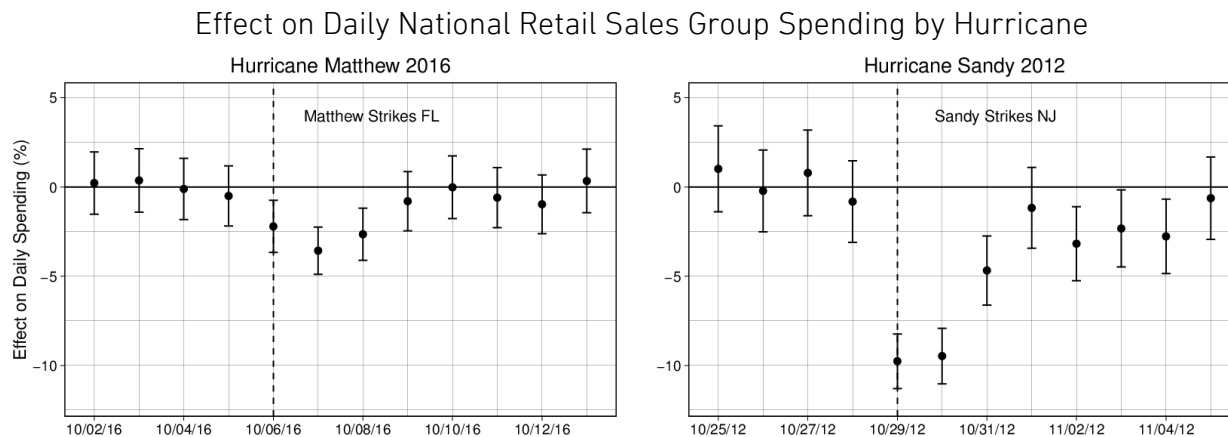
FIRST DATA/PALANTIR: HOW WELL DID IT WORK?

Filtered Index vs Census Retail Sales (12mo change, NSA)



CONSTRUCTING STATE-LEVEL INDEXES

A similar filtering algorithm was applied to **daily, state-level data** to construct daily indexes of spending within a given state. These are useful for analyzing things such as **effects of large storms on spending.**



Source: First Data Retail volume aggregates, Census Bureau.

Note: State-level spending indexes aggregated using 2012 Economic Census weights. The dots are the estimated coefficients for daily hurricane effect controlling for day of week, month, and trends. Whiskers are 95 percent confidence interval.

Additional details on this analysis are available in our recently published [FEDS Note](#).

SUMMARY AND FURTHER WORK

- Transaction aggregates can be useful sources of independent, timely signal for aggregate spending
- Requires care in constructing aggregate index
 - Transparency of transformations between data provider and user helps
- Additional work:
 - Current analysis of events, tax holidays, weather events, etc.
 - Gathering additional history to apply X13
 - Home-level data using geography of spending patterns