

Improving Regional PCE Estimates Using Credit Card Transactions Data

Abe Dunn
Ledya Guci
Mahsa Gholizadeh
Bryn Whitmire
Advisory Committee Update
July 20, 2016



The purpose of this presentation is to report on recent work using credit card transactions data to improve regional PCE estimates. The project uses data from a credit card intermediary called First Data and a tech firm called Palantir that specializes in analyzing and managing large databases.

Exploratory Work With First Data/Palantir

Data and coverage

- ~50% of all U.S. Credit Card transaction spend
- Data from 4.5M+ U.S. merchant locations
- 600+ merchant categories in our data set
- 58B transactions annually
- 1.6 Trillion spend, 10% of GDP
- All card-types, all banks, all networks, all 50 states, all customer segments, all merchant sizes
- 800M+ cardholders, 100% transactions from each merchant

This pilot uses customized aggregate data that includes:

- National estimates
- Flow of spending across geography by establishments and consumer location

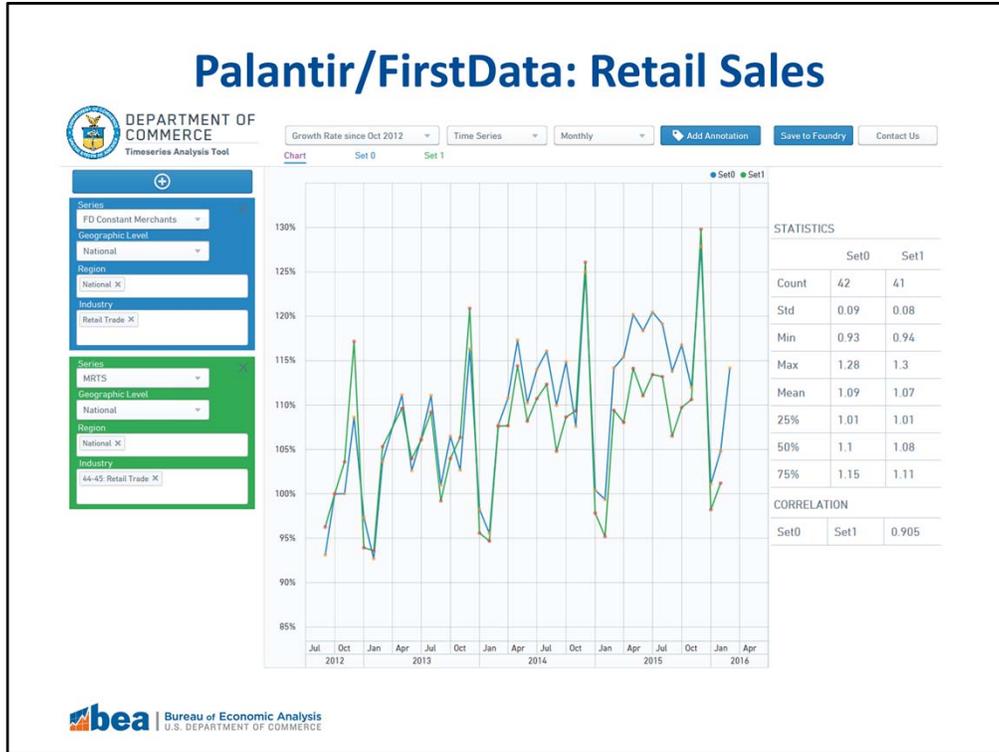
BEA has signed a contract with Palantir for a four-month term license to Palantir's commercial software, which will include access to data provided by First Data Corporation. First Data is an electronic card processor for about half of all commercial credit/debit card spending.

During the four-month term, Palantir will lend BEA the hardware necessary to deploy Palantir's commercial software, which will be installed within BEA's server room. There are 16 of their servers at BEA.

This pilot data set includes all transactions for 5 states (MA, IA, CA, TX and NY) for all industries coded using the MCC (merchant category code), from Aug 2012 to Nov 2015.

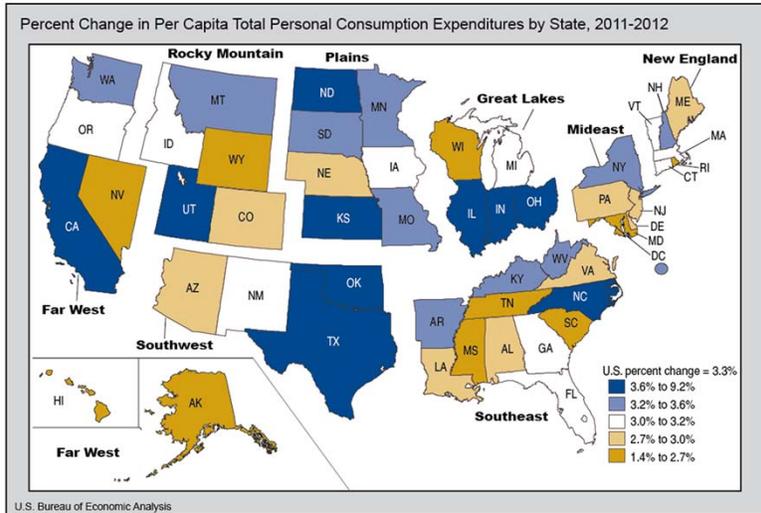
Data will be tabulated by month/county/industry and normalized to provide trends or shares. Palantir has also loaded the published PCE trend data (and PCE by state data) into the tool, and will load the retail trade and Economic Census public use data.

Additional flow of data spending for the entire U.S. has been provided. For instance, share of spending in each state and industry that is derived from individuals in other states.



Using the tool developed by Palantir, BEA is able to compare First Data estimates with other published surveys and statistics. For example, at the national level, here is the retail sales data (MMC 44-45) from the weighted First Data dataset (blue) compared to Census' Monthly Retail Trade Survey (green).

Help to Improve State-Level Estimates of Personal Consumption Expenditures (PCE) and May Help Generate MSA-Level Estimates



Given the relatively short time horizon of the data from 2012 to 2015, we have focused on using the data to help improve the accuracy and obtain better estimates in PCE at the state level or even at the MSA level.

A particular challenge for constructing PCE estimates at the regional level is that to construct the PCE estimates, we actually start with estimates of receipts at the level of the establishment from the Census. However, these estimates may be different than the consumption of individuals in that location if individuals consume in locations outside the location in which they reside. This poses a challenge at the state level and would be a greater issue for forming PCE estimates at the MSA level.

Adjusting Establishment Estimates from Census to Construct Regional PCE Statistics

Current process

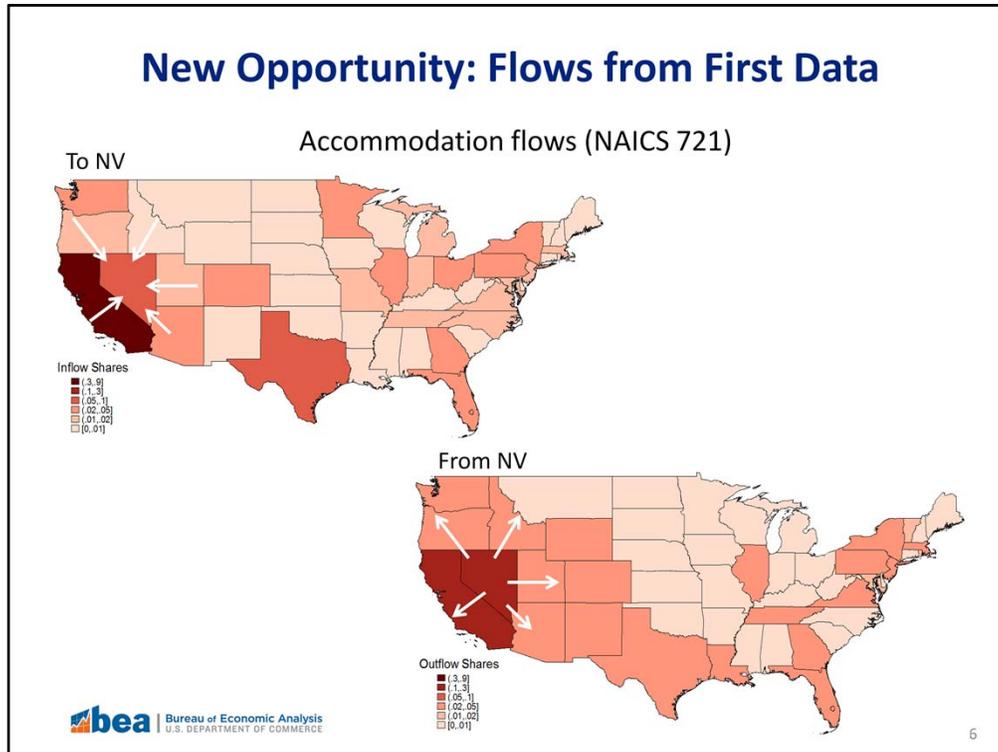
- Criteria for adjustment
 - Sufficient evidence of out-of-state spending
 - Economic reason for adjustment
 - A good category match available in consumer expenditure survey data
- Method
 - Adjust Census-based share with survey-based share
 - Rescale to national accounts totals

First Data spending flows allow for a new approach.

Our current methodology does attempt to make some adjustment for this border-crossing problem, although the adjustment is limited. We must rely on economic reasoning and data contained in the Consumer Expenditure Survey (CES). Basically, we are looking for strong evidence that the states are outliers based on the observed consumption and would be good candidates for adjustment. For instance, the large number of individuals that vacation in Nevada would make it a strong candidate for adjustment.

We use the First Data information on the locations of the establishment and the consumer to adjust for spending flows. In other words, rather than relying on judgment and limited information as a basis for adjustment, we are able to adjust all geographic areas automatically based on the consumption flow information.

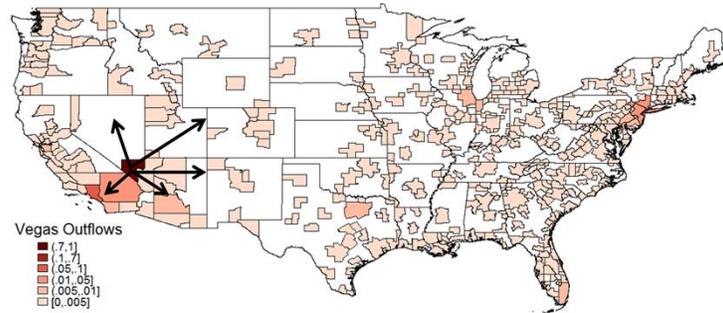
New Opportunity: Flows from First Data



These figures show the flow information for one example, accommodations in Nevada. We find that consumers in California, Texas, and Nevada have the largest shares of spending on accommodations within Nevada. Meanwhile, the state of Nevada is spending the most on accommodations within its own state and California.

Flow of Spending, Clothing and Footwear, 2012

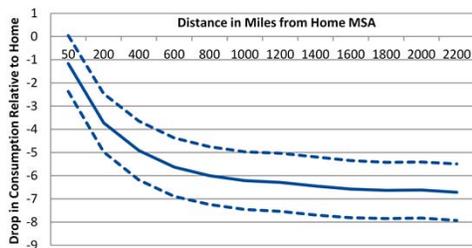
Consumption Outflow for Las Vegas MSA Residents



Taking it a step further and looking at the more disaggregate MSA flow data, we can look at spending trends on clothing and footwear for Las Vegas MSA residents. As expected, most consumption is seen within the Las Vegas MSA, but we do see increased activity in the Los Angeles MSA, other nearby geographic locations, and the New York City MSA.

Determinants of Spending Flows Using MSA-Level Flow Information, Clothing and Footwear

Consumption changes with distance from home



Destination location "quality", relative to home

MSA Name	"Quality"	Rank
Los Angeles-Long Beach-Anaheim, CA	5.83	1
New York-Newark-Jersey City, NY-NJ-PA	5.43	2
Las Vegas-Henderson-Paradise, NV	5.38	3
San Francisco-Oakland-Hayward, CA	5.24	4
Miami-Fort Lauderdale-West Palm Beach, FL	4.90	5
Seattle-Tacoma-Bellevue, WA	4.54	6
Dallas-Fort Worth-Arlington, TX	4.49	7
Chicago-Naperville-Elgin, IL-IN-WI	4.21	8
Phoenix-Mesa-Scottsdale, AZ	4.20	9
San Diego-Carlsbad, CA	4.17	10
Orlando-Kissimmee-Sanford, FL	4.12	11
Washington-Arlington-Alexandria, DC-VA-M	3.94	12
Philadelphia-Camden-Wilmington, PA-NJ-DE	3.89	13
Riverside-San Bernardino-Ontario, CA	3.85	14
Boston-Cambridge-Newton, MA-NH	3.84	15
Urban Honolulu, HI	3.81	16
Portland-Vancouver-Hillsboro, OR-WA	3.76	17
Atlanta-Sandy Springs-Roswell, GA	3.62	18
Tampa-St. Petersburg-Clearwater, FL	3.51	19
St. Louis, MO-IL	3.46	20

bea Bureau of Economic Analysis
U.S. DEPARTMENT OF COMMERCE

8

Before applying these data and to better understand patterns contained in these data, we estimate a simple model of consumption flow in which individuals choose where to consume. The key explanatory variable is share of consumption in an MSA and the determinants of these consumption flows are distance from an individual's home MSA and MSA fixed effects.

As expected, we find that the greater the distance from the home location, the less likely the consumer is to spend at that location.

Ranking the MSA's by "most desirable" places to spend, we find the Los Angeles MSA at the top of the list followed by NYC and Las Vegas. As expected, we see many large cities and top vacation spots at the top of this list.

After checking these basic patterns in the data, we next use these data to correct for regional PCE estimates. To implement the correction, we essentially use the flow information to send consumption from the location of the establishment, as reported in the economic Census, to the location of the consumer.

Consumption Flows and State-Level PCE Estimates



Preliminary Estimates

Food Services and Accommodations, 2012

Initial estimates		Difference from U.S. Value	Per Dollar of Disposable Income	Difference from U.S. Value
Geography	Per Capita			
United States	\$2,181	0.0%	0.055	0.0%
Illinois	\$2,193	0.6%	0.054	-2.2%
Hawaii	\$5,807	166.2%	0.144	159.9%
Nevada	\$3,992	83.0%	0.111	101.5%

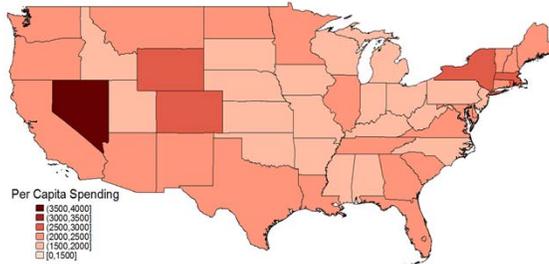
Incorporating FD flows		Difference from U.S. Value	Per Dollar of Disposable Income	Difference from U.S. Value
Geography	Per Capita			
United States	\$2,181	0.0%	0.055	0.0%
Illinois	\$2,359	8.2%	0.058	5.2%
Hawaii	\$2,763	26.7%	0.068	23.7%
Nevada	\$1,578	-27.7%	0.044	-20.3%

Next, we apply the flow information to the data. This table demonstrates the influence of using First Data flow information for combined MCC for food service and accommodations for select states. The table in the top of the slide shows the initial estimates based on Census data. These estimates show per capita spending values based on establishment receipt information. Hawaii and Nevada are clear outliers based on several criteria.

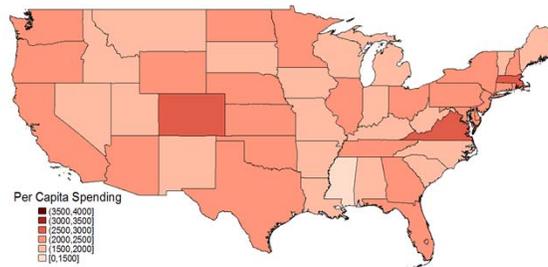
The bottom table shows what happens when First Data flow information is used to automate the adjustment. The use of First Data flows automatically corrects outlier problems. The change can be seen geographically in the next slide.

Food Services and Accommodations, 2012

Initial estimates



Incorporating FD flows



 Bureau of Economic Analysis
U.S. DEPARTMENT OF COMMERCE

11

Here we show the nationwide impact of the before and after adjustment on per capita spending on food services and accommodations. High spending areas such as New York and Nevada are no longer outliers after the adjustment.

Preliminary Estimates

Clothing and Footwear, 2012

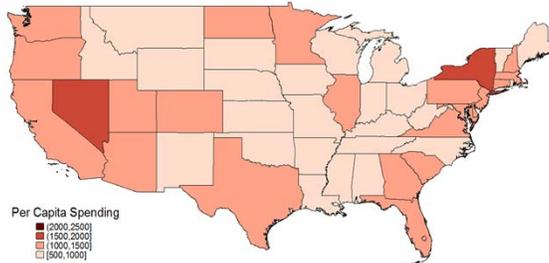
Initial estimates		Difference from U.S. Value	Per Dollar of Disposable Income	Difference from U.S. Value
Geography	Per Capita			
United States	\$1,128	0.0%	0.029	0.0%
Illinois	\$1,154	2.3%	0.028	-0.5%
Hawaii	\$1,813	60.8%	0.045	57.0%
Nevada	\$1,796	59.3%	0.050	75.4%

Incorporating FD flows		Difference from U.S. Value	Per Dollar of Disposable Income	Difference from U.S. Value
Geography	Per Capita			
United States	\$1,128	0.0%	0.029	0.0%
Illinois	\$1,149	1.8%	0.028	-0.9%
Hawaii	\$1,339	18.7%	0.033	15.9%
Nevada	\$1,071	-5.1%	0.030	4.5%

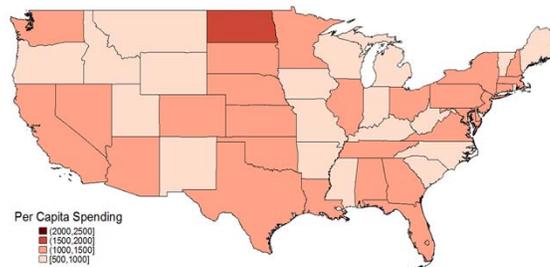
Same exercise is repeated but using a different industry. Very similar results are obtained.

Clothing and Footwear, 2012

Initial estimates



Incorporating FD flows



bea Bureau of Economic Analysis
U.S. DEPARTMENT OF COMMERCE

13

Here we show the nationwide impact of the before and after adjustment on per capita spending on clothing and footwear.

Consumption Flows and MSA-Level PCE Estimates



Preliminary Estimates

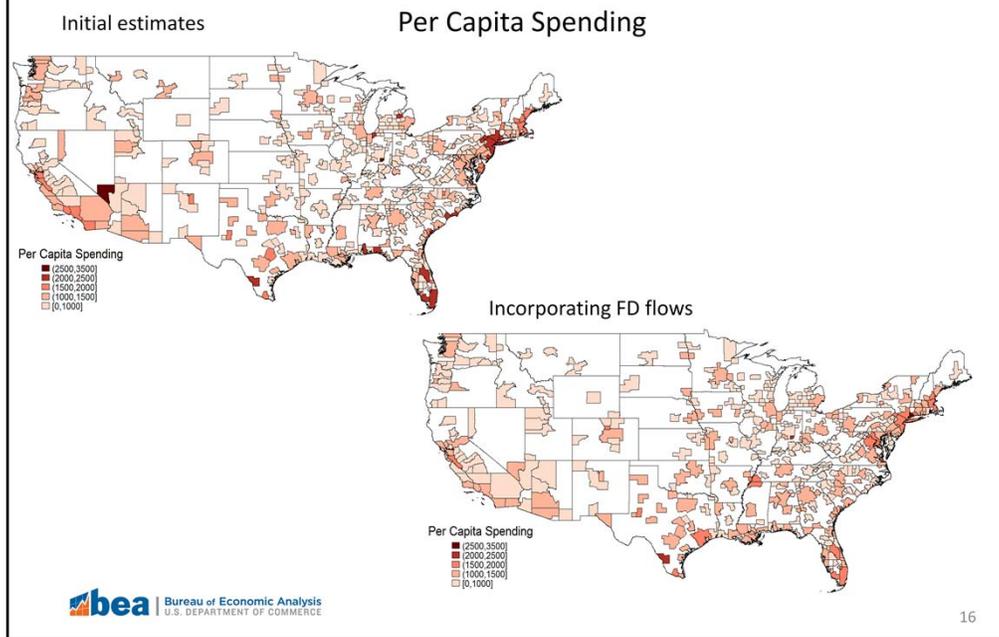
Clothing and Footwear, 2012

Initial estimates		Difference from U.S. Value	Per Dollar of Personal Income	Difference from U.S. Value
Geography	Per Capita			
United States	\$1,128	0.0%	0.026	0.0%
Kansas City, MO-KS	\$995	-11.7%	0.022	-12.8%
Kahului-Wailuku-Lahaina, HI	\$2,683	137.9%	0.070	175.3%
Las Vegas-Henderson-Paradise, NV	\$2,936	160.3%	0.076	197.7%

Incorporating FD flows		Difference from U.S. Value	Per Dollar of Personal Income	Difference from U.S. Value
Geography	Per Capita			
United States	\$1,128	0.0%	0.026	0.0%
Kansas City, MO-KS	\$1,094	-3.0%	0.024	-4.2%
Kahului-Wailuku-Lahaina, HI	\$1,174	4.1%	0.031	20.5%
Las Vegas-Henderson-Paradise, NV	\$1,163	3.1%	0.030	17.9%

Again, this example is similar to the others above, but uses MSA estimates rather than state estimates. Similar results are found. These estimates suggest that MSA-level PCE estimates are feasible.

Clothing and Footwear, 2012



Same as previous figures, but shows correction for MSA level data rather than state level data.

Spending Flows for PCE by State

Opportunities

- Flow shares can be readily incorporated and simplify the current methodology
- Spending and consumption flows across areas provide a unique view of geography of consumption

Considerations

- Varying data quality and coverage by industry and by geography
- Imputation of consumer location

Using the First Data flow data would provide BEA the opportunity to improve current methodology for calculating PCE by state, but we may also be able to provide unique consumption flow information. Consumption flow information may be particularly useful for policymakers and businesses trying to better understand the interactions among regional economies.

There is still a considerable amount of work to be done. For instance, we need to be careful about applying these methods across all industries and geographies, as coverage of establishment varies by geography. A second important consideration is that we do not have the precise location of all the individuals in the data. Therefore, for many individuals an algorithm must be applied to guess the consumer's location. More work is necessary to improve this algorithm.

Next Steps

- Refine the home location algorithm and further evaluate flow information.
- Investigate e-commerce data.
- Continue working with data to refine adjustment for regional PCE.

We are continuing to refine our home location algorithm. The quality of the estimates greatly depends on the accuracy of this home location algorithm.

The pilot project does not include e-commerce information. In our pilot study, we would also like to expand our analysis to include e-commerce.

These slides have presented a relatively simple and stylized application of how this flow correction would work. In forming the actual regional PCE estimates, there are many additional steps involved. Our future work would examine how to incorporate this flow information in practice.

Additional Slides



Preliminary Estimates

Food Services and Accommodations, 2012

Initial estimates		Difference from U.S. Value	Per Dollar of Personal Income	Difference from U.S. Value
Geography	Per Capita			
United States	\$2,181	0.0%	0.049	0.0%
Kansas City, MO-KS	\$2,117	-3.0%	0.047	-4.2%
Kahului-Wailuku-Lahaina, HI	\$9,597	340.0%	0.251	409.3%
Las Vegas-Henderson-Paradise, NV	\$7,707	253.4%	0.199	304.1%

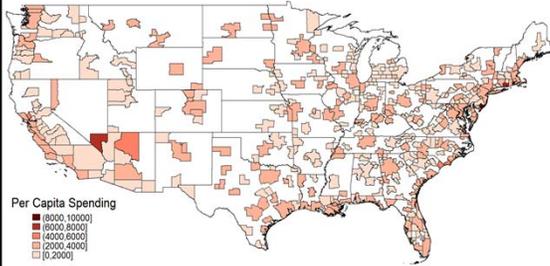
Incorporating FD flows		Difference from U.S. Value	Per Dollar of Personal Income	Difference from U.S. Value
Geography	Per Capita			
United States	\$2,181	0.0%	0.049	0.0%
Kansas City, MO-KS	\$2,317	6.2%	0.052	4.9%
Kahului-Wailuku-Lahaina, HI	\$1,804	-17.3%	0.047	-4.2%
Las Vegas-Henderson-Paradise, NV	\$1,957	-10.3%	0.051	2.6%

Similar to above estimates, but for a different industry.

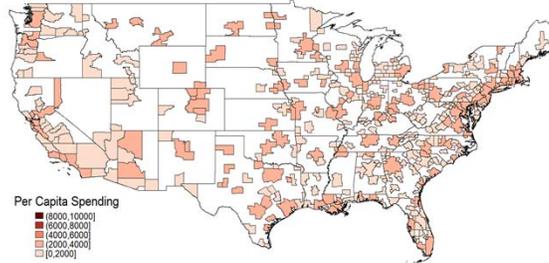
Food Services and Accommodations, 2012

Initial estimates

Per Capita Spending

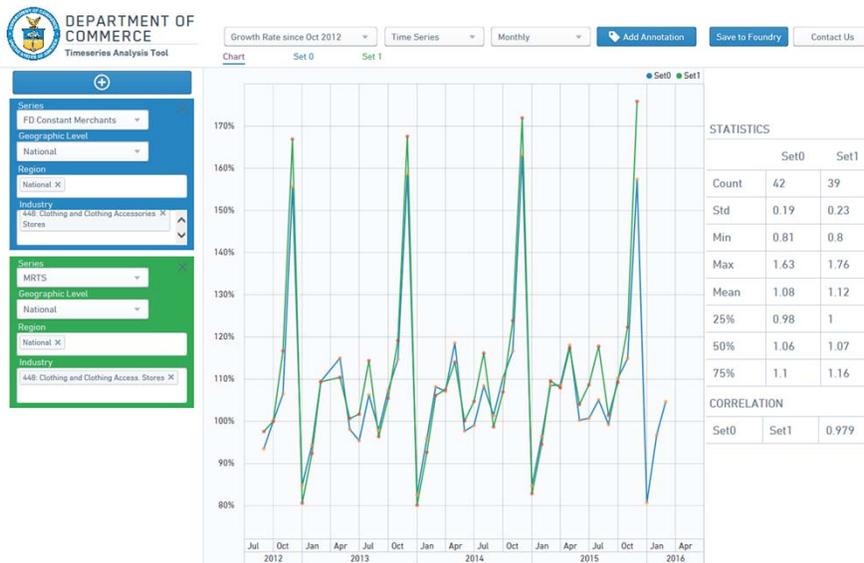


Incorporating FD flows



 **Bureau of Economic Analysis**
U.S. DEPARTMENT OF COMMERCE

Palantir/FirstData - 448: Clothing Stores



bea Bureau of Economic Analysis
U.S. DEPARTMENT OF COMMERCE

The match in retail sales and is particularly close for some categories, such as clothing. This slide demonstrates that the match with official retail sales greatly depends on the industry that one is looking at.