

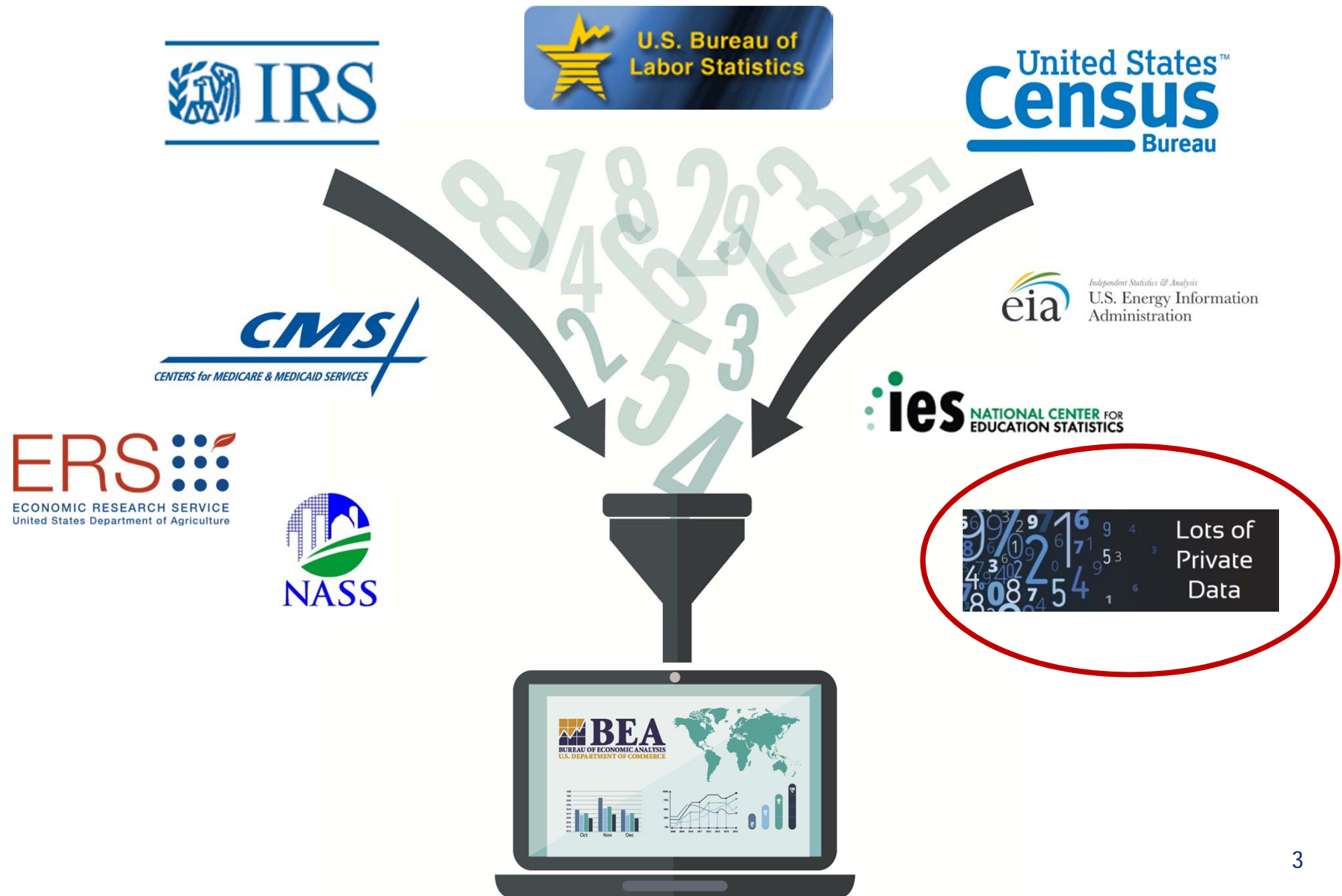
Exploiting Commercial Data for Official Economic Statistics

David Johnson
CNSTAT expert meeting
November 19, 2015

Motivation and Goals

- Improve early estimates
- More detailed regional data
- Passive data collection

BEA is a Data Aggregator



Using Big Data—not a big change for BEA

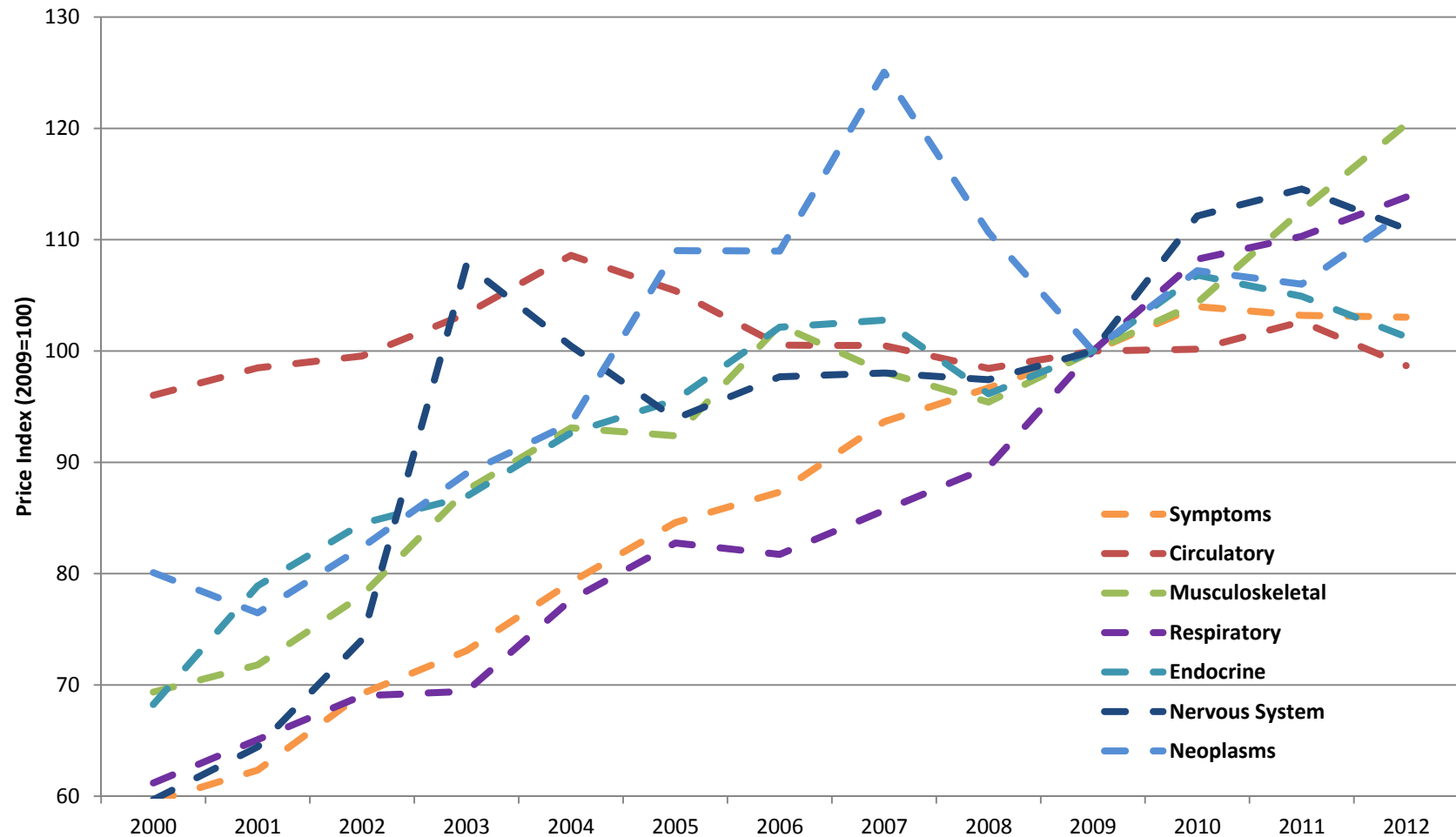
- Examples of private data source used by BEA; 121 sources, over \$1million spent

Source Data	Industry	International	National	Regional
AM Best	X		X	X
Bureau Van Dijk		X		
Compustat	X		X	
Mercer's Health Plan				X
Merger Market		X		
Pharma	X		X	
RL Polk			X	
Ward's Automotive			X	

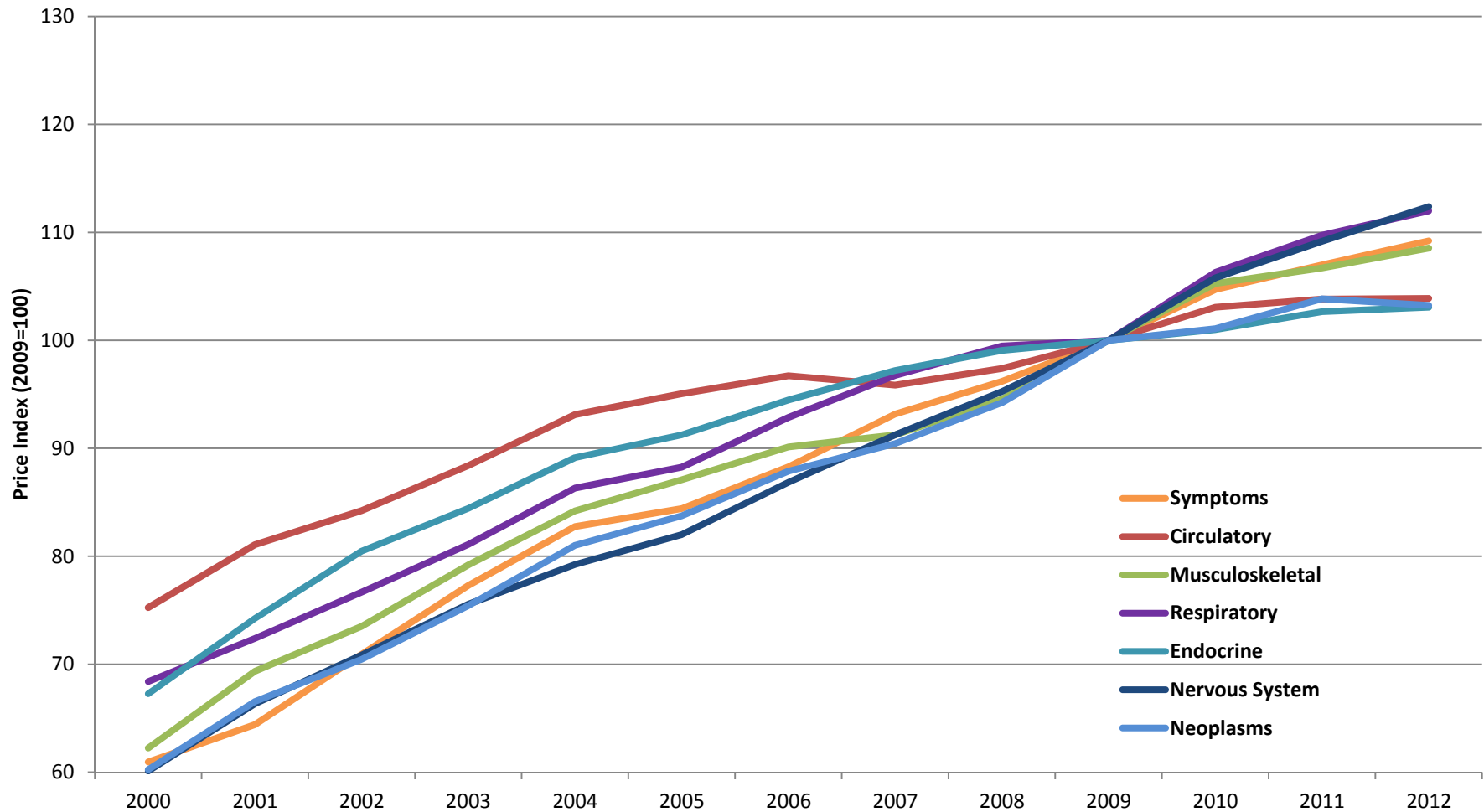
Example: Satellite Health Care Account

- Data come from many sources and are blended
- BEA combined billions of claims from both Medicare and private commercial insurance to determine the spending for over 250 diseases

Example: Health Care Satellite Account Prices--Survey Data Only



Example: Health Care Satellite Account Prices--Survey + Big Data



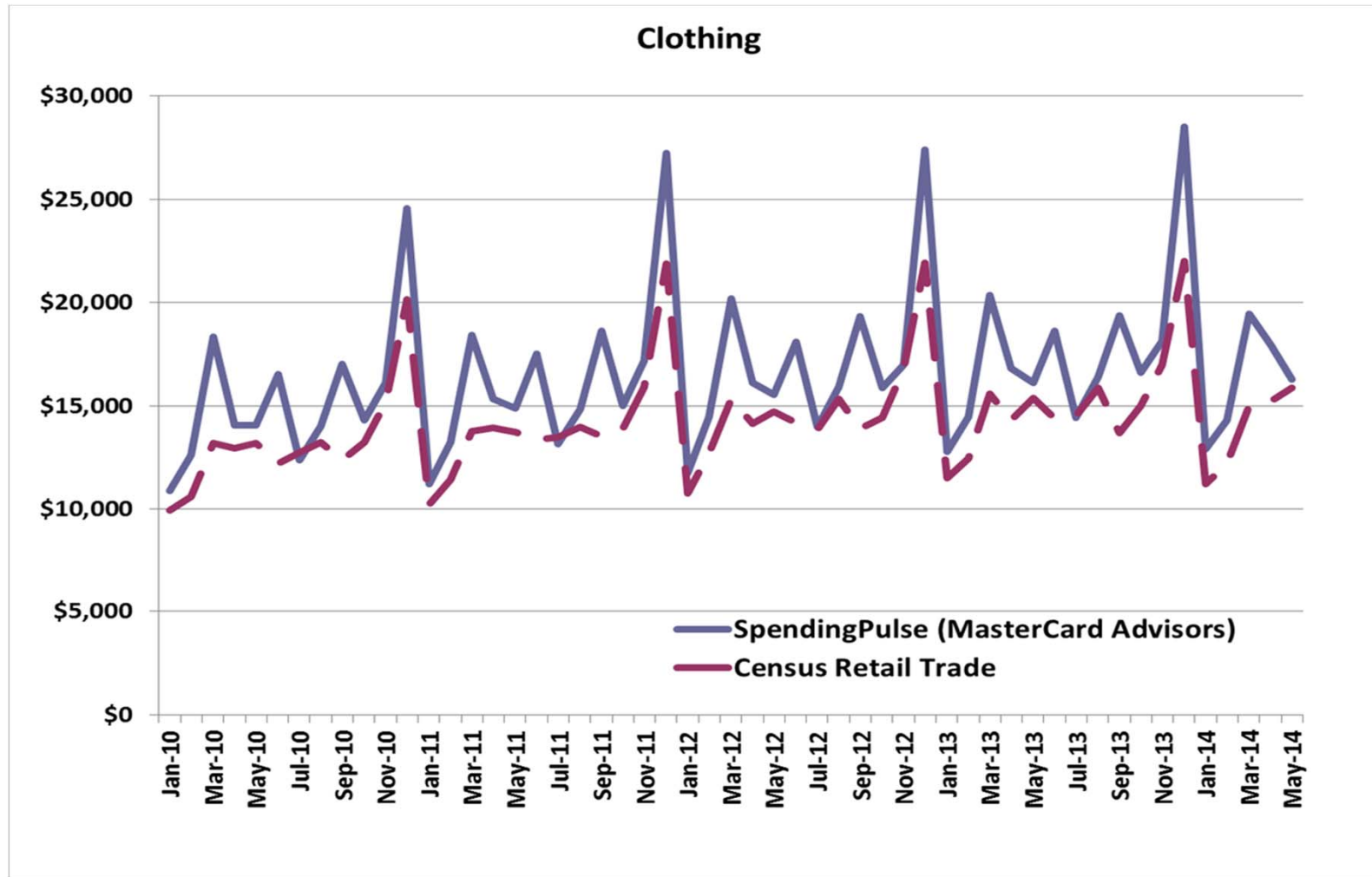
Exploring Athenahealth Data

- Data from Athenahealth provides a convenience sample of physicians offices with information on dollars received by providers.
- Importantly the data are available in “real time.”
- The goal of this project is to assess whether these data could be used to improve the advance and 2nd estimates for nominal spending on physician services, which uses QSS.

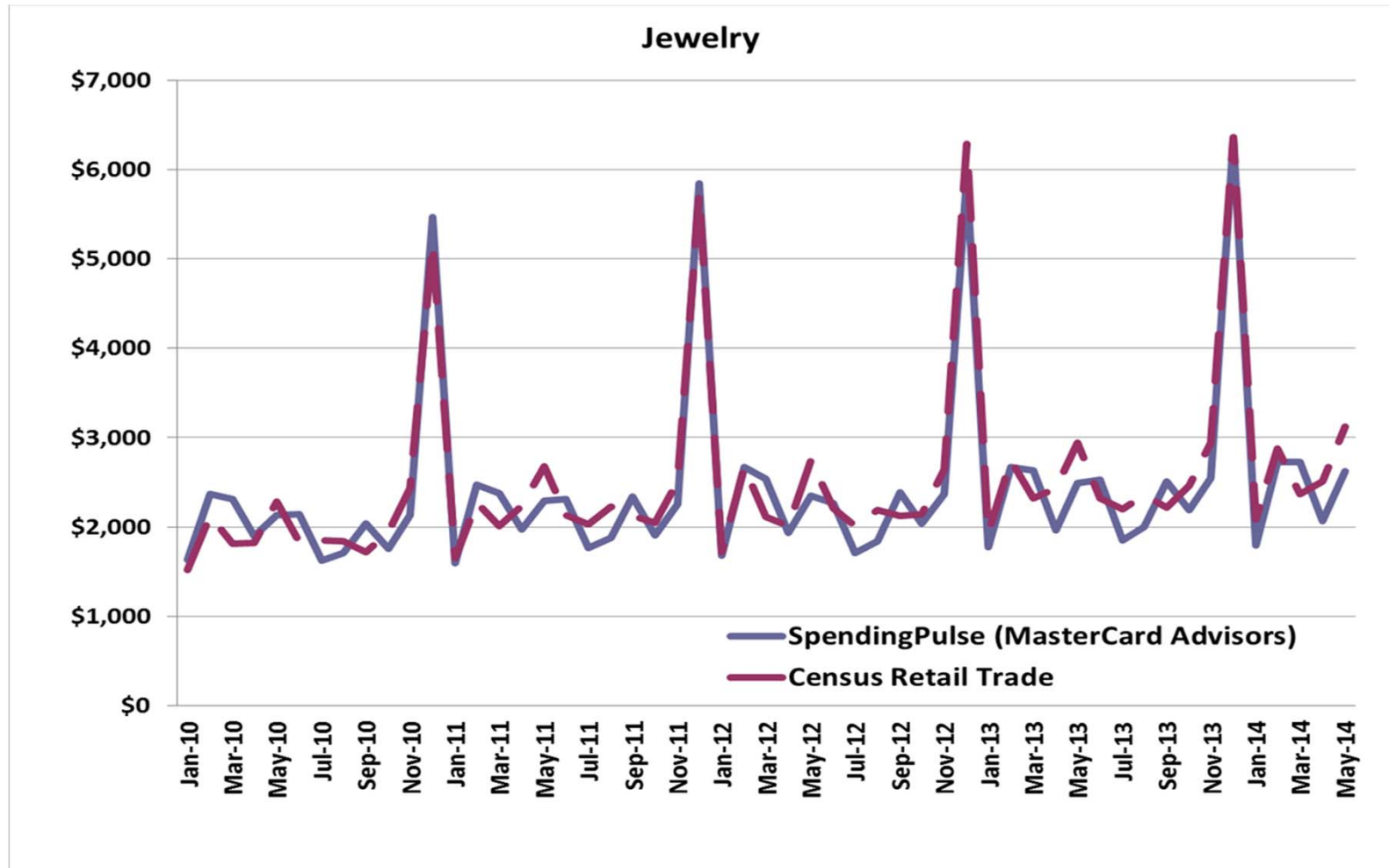
Credit Card Data for Consumer Spending

- Exploring use of credit card data to improve estimates of consumer spending, and to develop estimates at the metro area and county levels
- Working with Census Bureau on acquiring and analyzing the data.
- Pilot Projects
 - MasterCard
 - First Data/Palantir
- Exploring Nielsen, PayPal, JPMorgan Chase

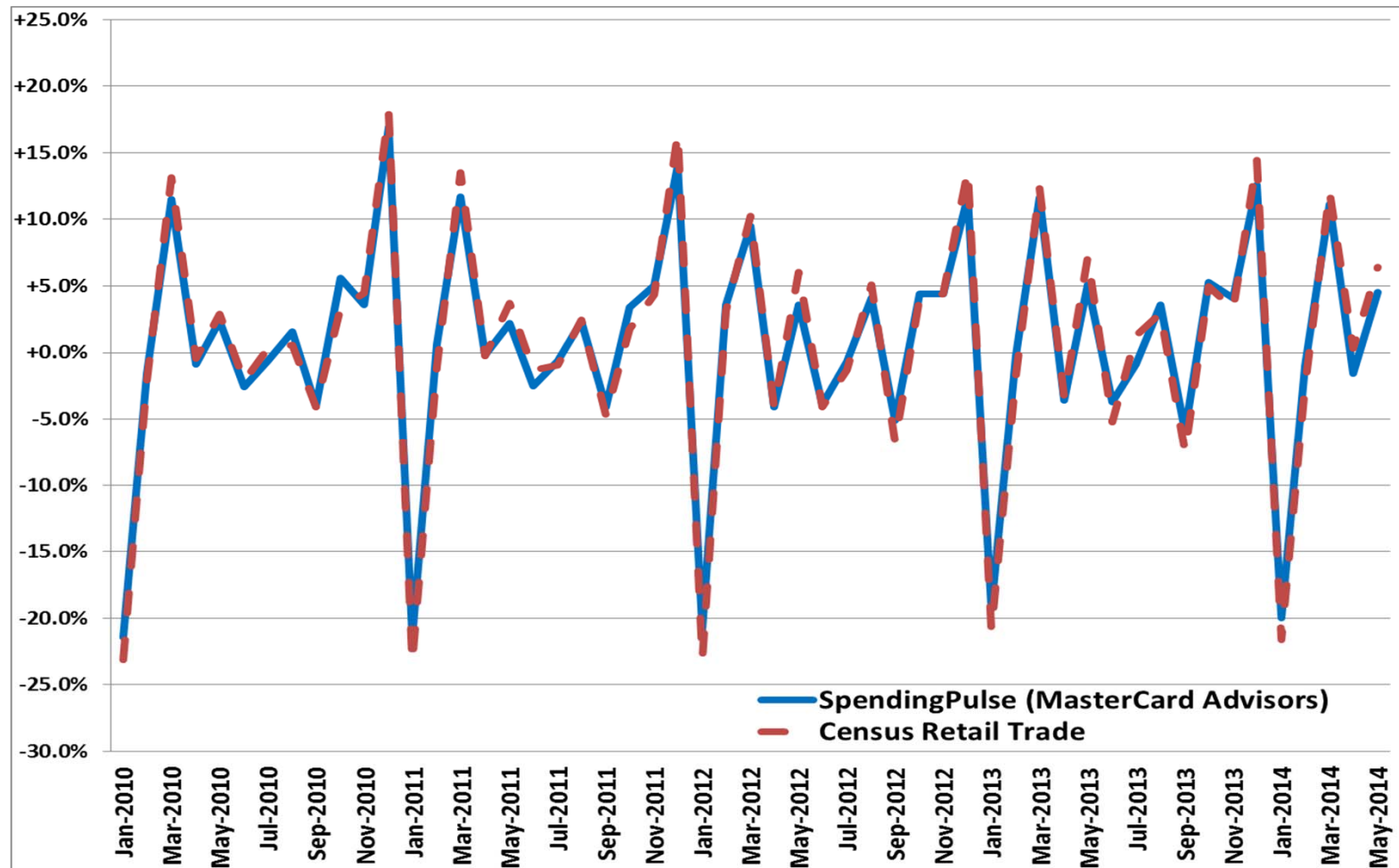
Monthly credit card data estimates compared to retail trade estimates



Monthly credit card data estimates compared to retail trade estimates



Monthly changes in credit card and total retail trade (less autos)



Palantir/First Data tool can show shares and time trends by state/county and industry

DEPARTMENT OF
COMMERCE

Transaction Analysis Dashboard

Metric

Regional Share

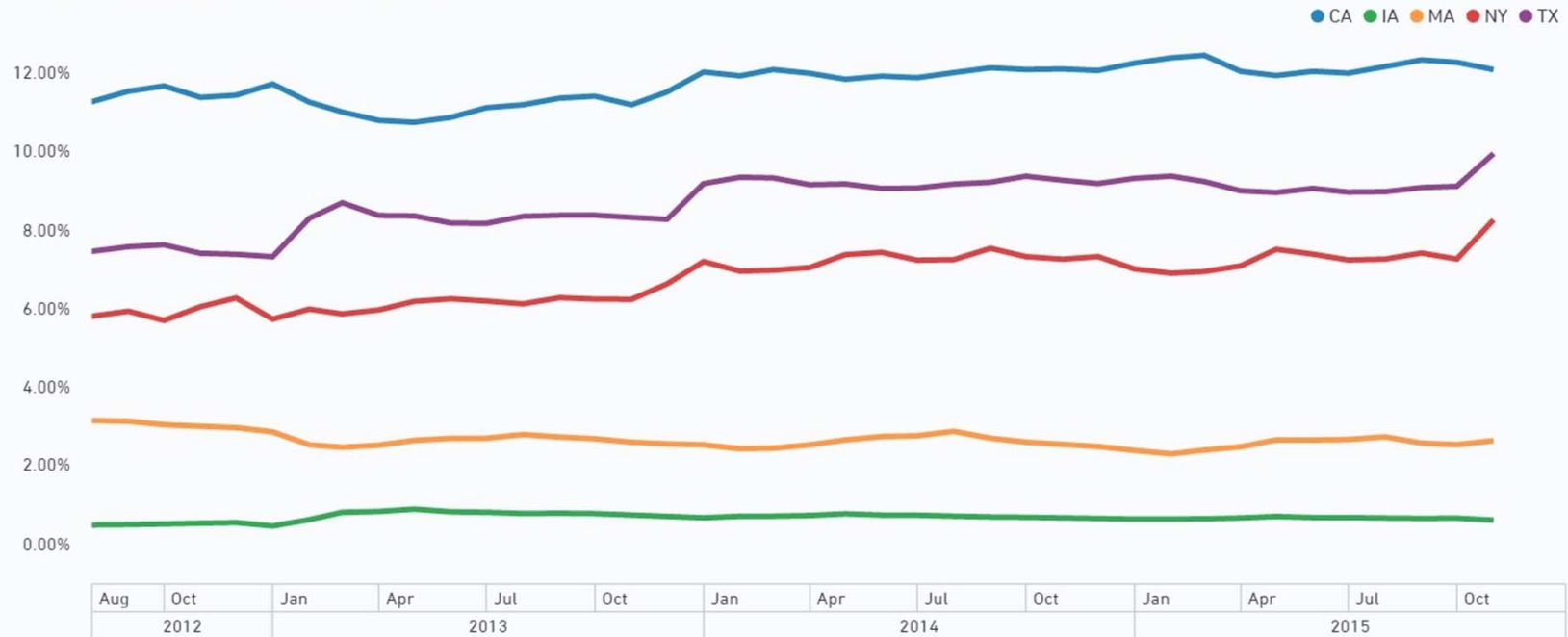
Region

State

Industry

All

STATE SHARE OF NATIONAL SALES



Gasoline share of total by State

DEPARTMENT OF
COMMERCE

Transaction Analysis Dashboard

Metric

Industry Share

Region

State

Industry

Gasoline and Other Energy ...

INDUSTRY SHARE OF TOTAL RETAIL SALES BY STATE

CA IA MA NY TX

30.00%
25.00%
20.00%
15.00%
10.00%
5.00%
0.00%

Aug	Oct	Jan	Apr	Jul	Oct	Jan	Apr	Jul	Oct	Jan	Apr	Jul	Oct
2012		2013				2014				2015			

Trend in gasoline sales by State

DEPARTMENT OF
COMMERCE

Transaction Analysis Dashboard

Metric

Growth Rate since August ...

Region

State

Industry

Gasoline and Other Energy ...

STATE GROWTH OF TOTAL SALES SINCE AUGUST 2012 FOR GASOLINE AND OTHER ENERGY GOODS



Compare the state trends to PCE trends (and can compare to PCE by state)

DEPARTMENT OF
COMMERCE

Transaction Analysis Dashboard

Metric

Growth Rate since August ...

Region

State

Industry

Gasoline and Other Energy ...

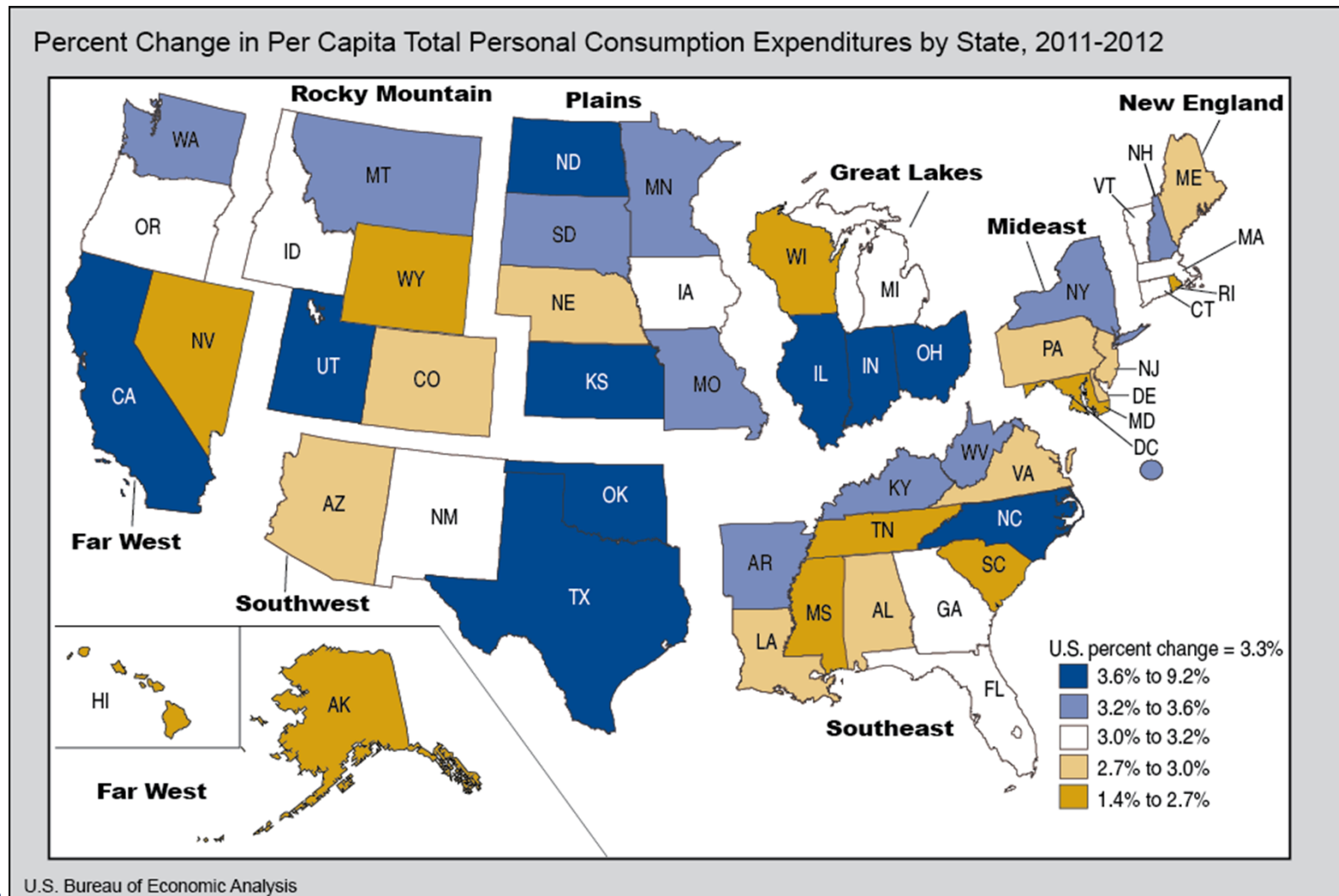
STATE GROWTH OF TOTAL SALES SINCE AUGUST 2012 FOR GASOLINE AND OTHER ENERGY GOODS



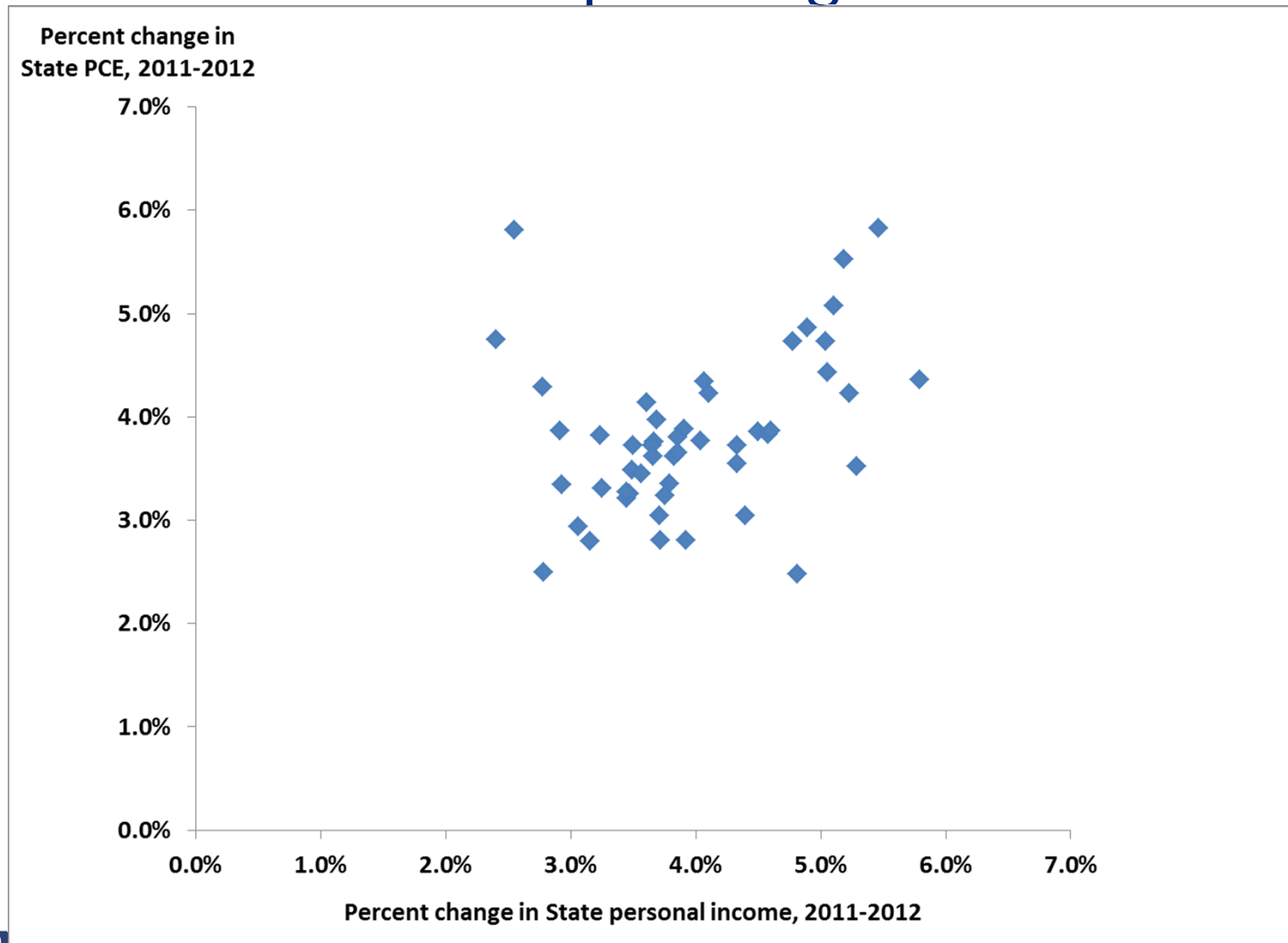
COMPARISON DATA SET

National PCE by Month

Use Commercial data to improve State level estimates of PCE between Economic Censuses, and may help generate county level estimates



Key is to link change in income to change in spending



Meeting Agenda

- Overview of challenges, opportunities and limitations
 - Trivellore Raghunathan and Simon Wilkie
- Reflections and possibilities from private sector
 - JPMorgan Chase Institute, MasterCard, Google, Zillow, Palantir/First Data
- Applications using commercial data
 - Jonathan Levin and Matthew Shapiro
- Discussion topics
 - projects in other countries
 - new technology
 - statistical issues
 - privacy issues
 - public/private partnerships

Questions for Discussion

- How could these data be used in official economic statistics?
 - For improving advance estimates
 - For improving regional estimates
 - For replacing survey data
- How do the estimates from these data compare to official estimates?
- What are the challenges in allowing agencies to access and use these data?