Big Data Landscape, 2015-2020

Brian C. Moyer, Director

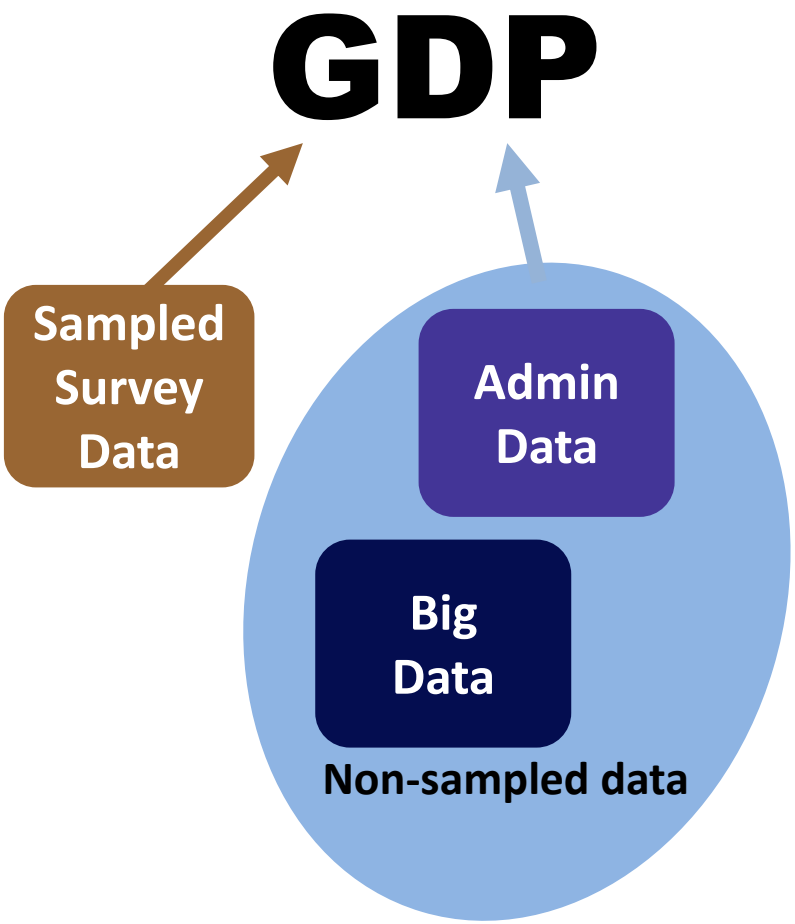
Global Conference on Big Data for Official Statistics

October 20, 2015

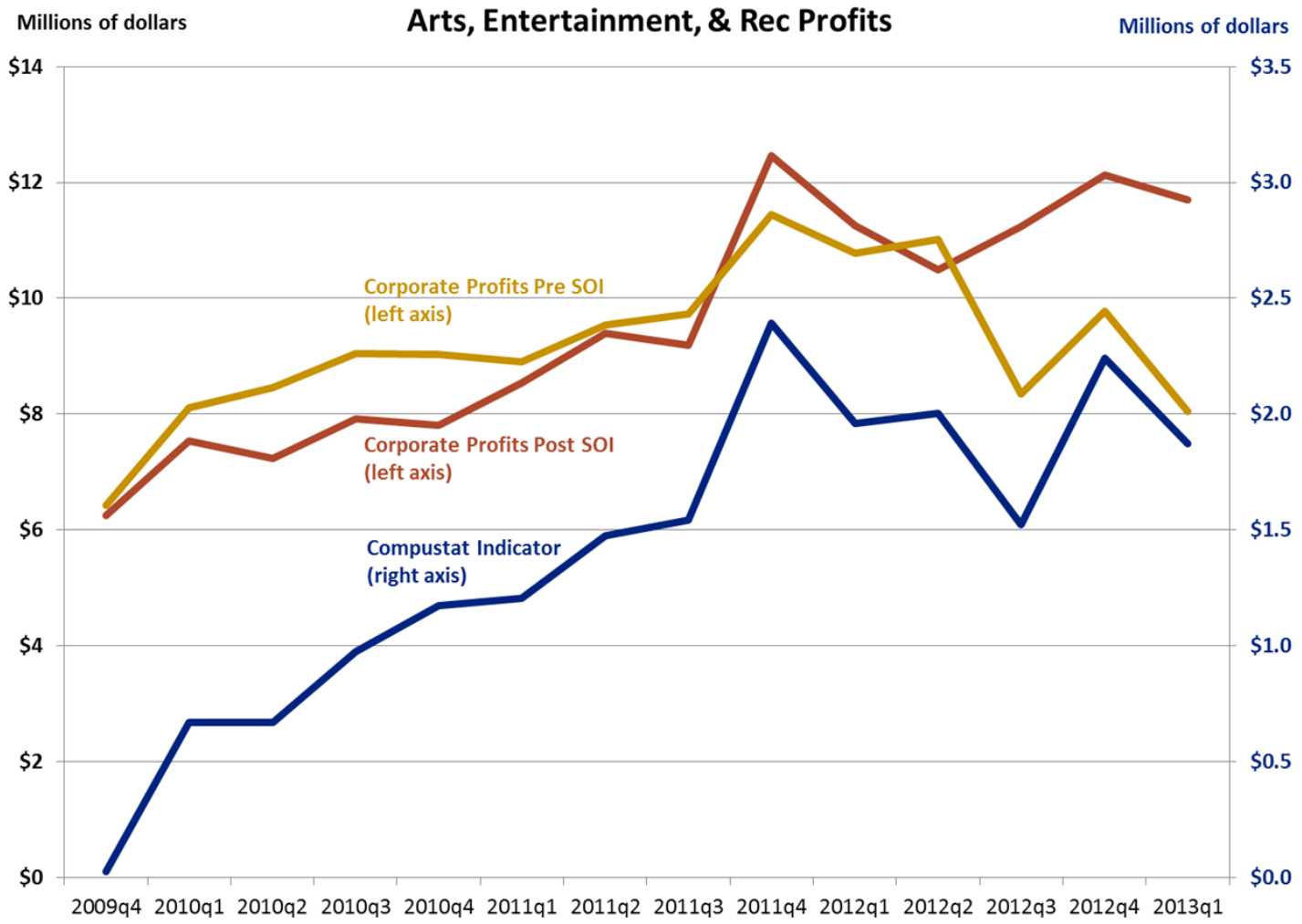
BEA
BUREAU OF ECONOMIC ANALYSIS
U.S. DEPARTMENT OF COMMERCE
www.bea.gov @BEA_News

Where are we now?

- Official statistics use a variety of big data sources
- Current private source data for GDP estimation:
  - Ward's/Polk/JD Power (auto sales/price/registrations)
  - American Petroleum Institute (oil drilling)
  - *Variety* magazine (motion picture admissions)
  - IRS, Statistics of Income
  - DOL, Unemployment Insurance data
- Source data for Producer Price Index estimation (Bureau of Labor Statistics):
  - Stock exchange security trades
  - Medicare Part B reimbursements



U.S. Corporate Profits



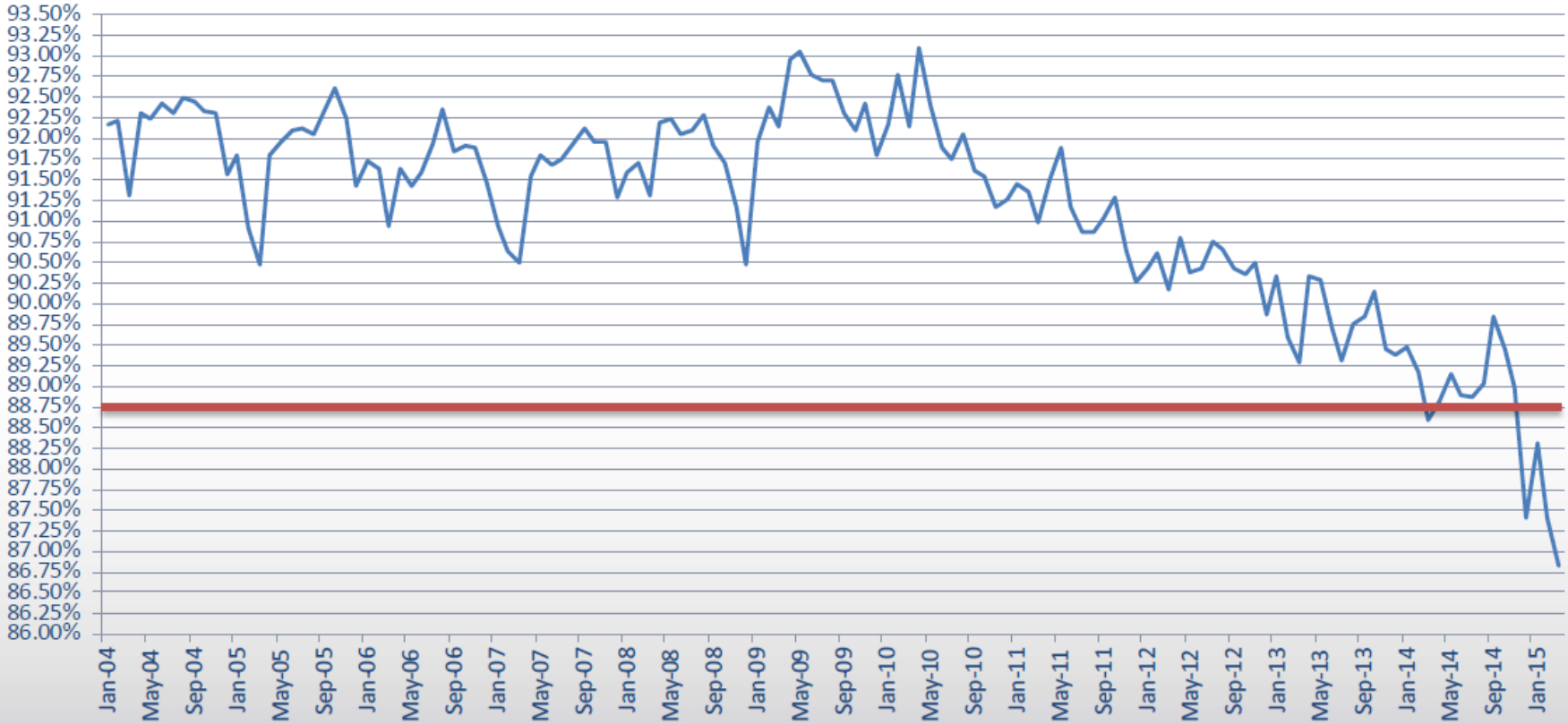
Some challenges with using big data

- How representative are the data?
- Do the concepts match those needed to measure output, prices, employment, etc.?
- Do the data provide consistent time series and classifications?
- Is it possible to bridge gaps in coverage?
- How timely are the data?
- How cost effective?
- What confidential issues arise and how limiting are they?

Are big data the answer?

- Need to continue efforts to explore Big Data as the landscape facing government agencies has changed:
  - Shrinking budgets
  - Increasing demand for more timely and detailed data
  - Increasing respondent burden
  - Emerging providers of economic statistics
  - Declining response rates

CPS-Response Rates by Month: January 2004 to Present

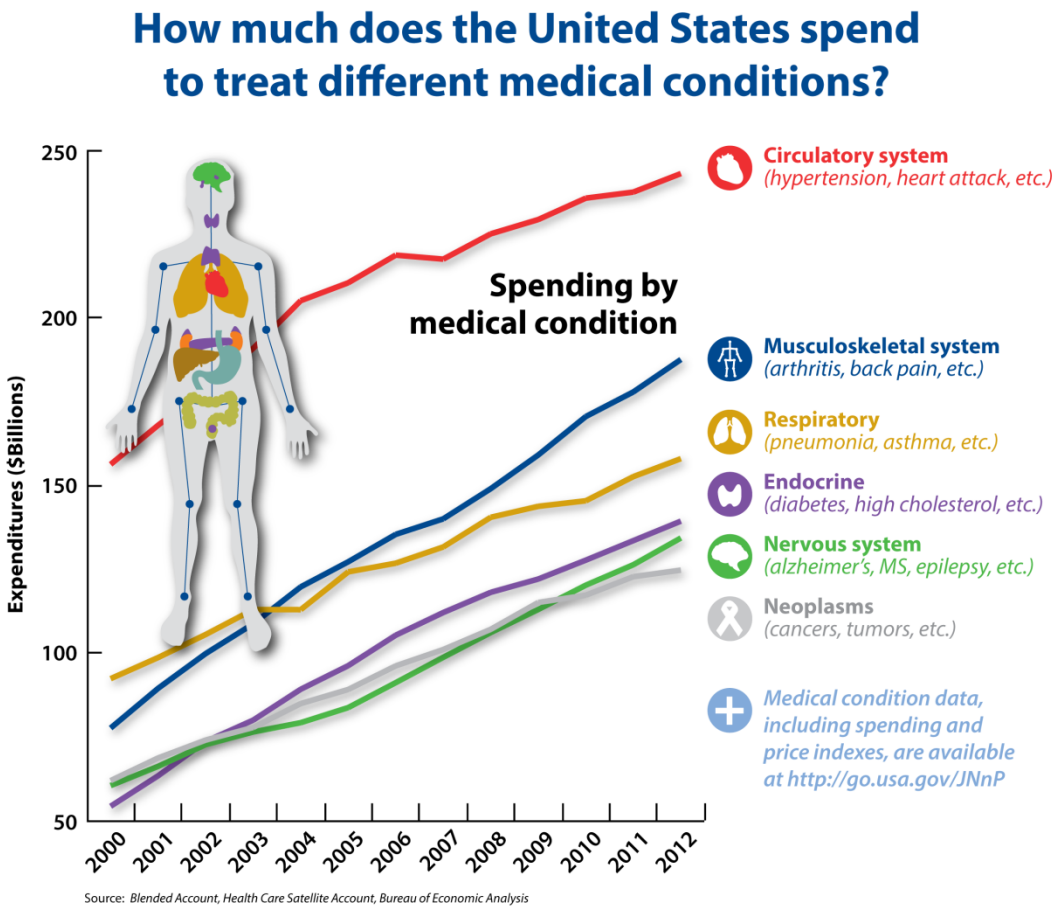


Where we are headed?

- Looking to big data to
  - **Extend**: provide more timely data, fill real time data gaps, expand geographical detail
  - **Enhance**: provide detailed interpolators, provide characteristic information for product quality, provide paradata for responsive survey design
  - **Verify**: confirm trends, validate findings from direct survey collection
  - **Supplement**: facilitate passive data collection

Recent Experience & Lessons Learned

Health Care Satellite Account



- Annual statistics for 2000-2012 that provide information on spending and price changes by disease category
- BEA combined billions of claims from both Medicare and private commercial insurance to determine the spending for over 250 diseases

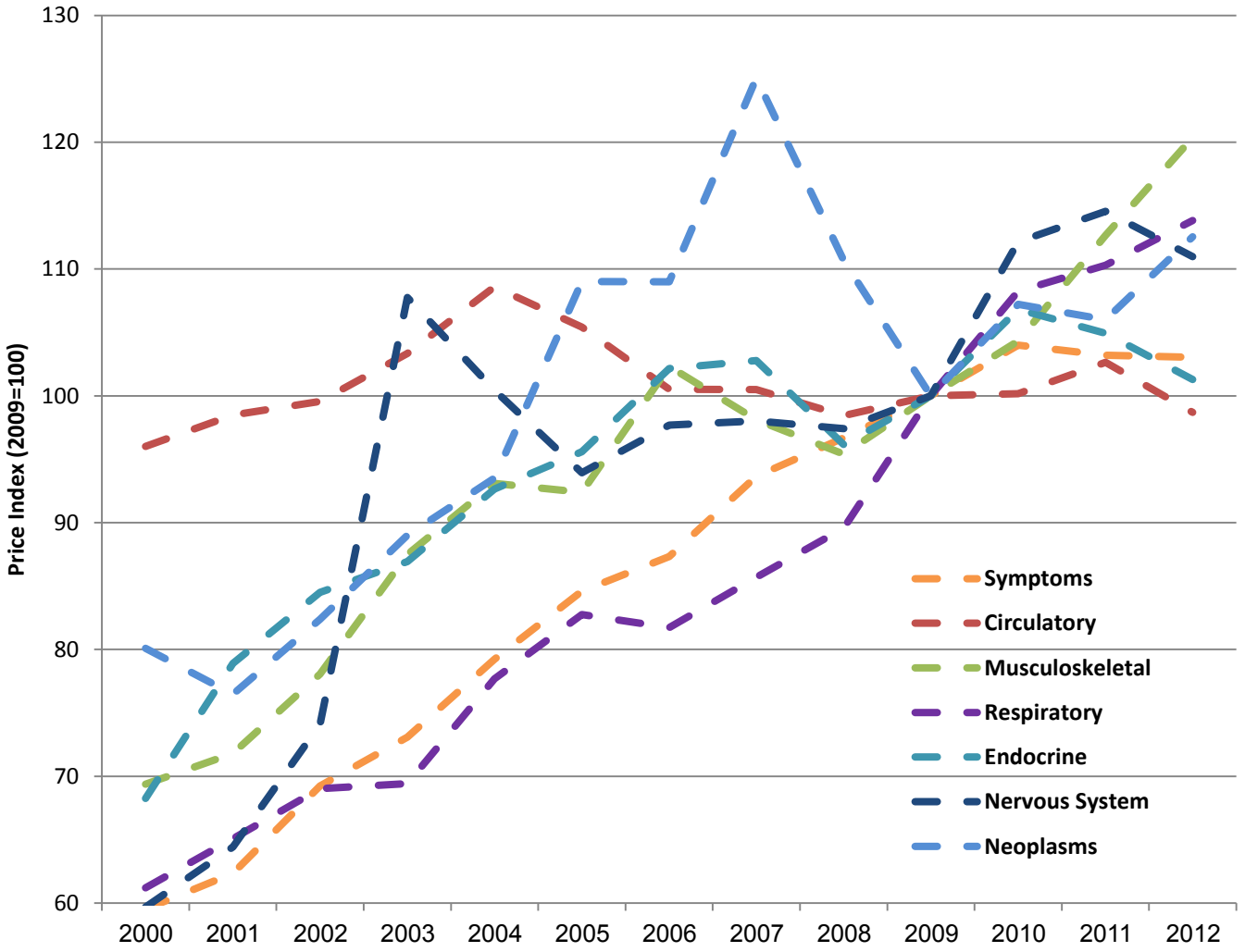
BEA BUREAU OF ECONOMIC ANALYSIS U.S. DEPARTMENT OF COMMERCE

For more information, refer to http://bea.gov/national/health_care_satellite_account.htm

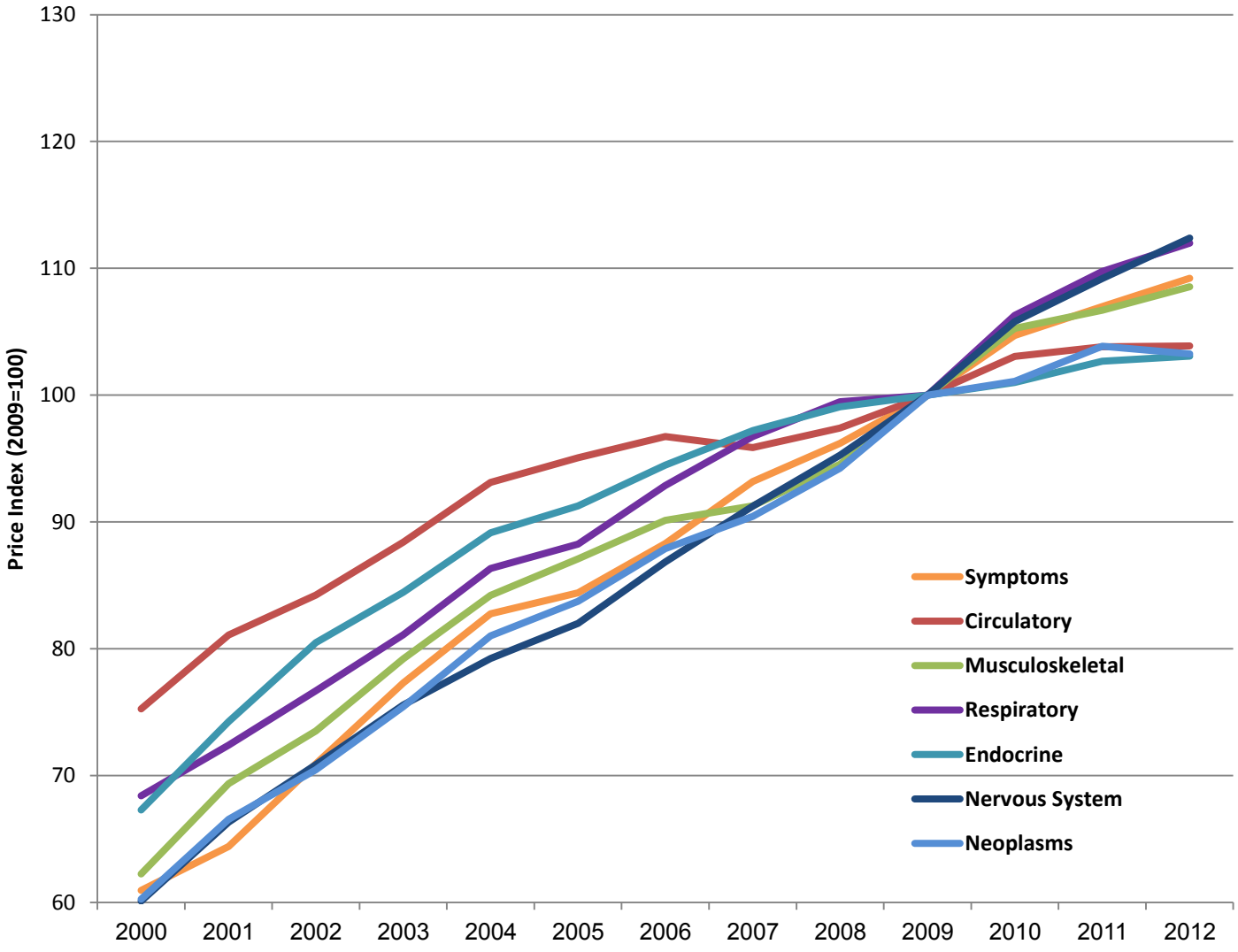
Health Care Satellite Account

- Two approaches:
  - "MEPS Account"
    - Using Medical Expenditure Panel Survey (MEPS)
    - Nationally representative sample with approximately 30,000 individuals
  - "Blended Account"
    - Includes MEPS data, private and public claims data
    - Commercially-insured patients from the MarketScan® Data from Truven Health
      - More than 2 million enrollees in each year
      - Convenience sample → application of population weights
    - Medicare patients from 5 percent random sample of enrollees
      - Approximately 2 million enrollees each year

Health Care Satellite Account: Survey Data Only



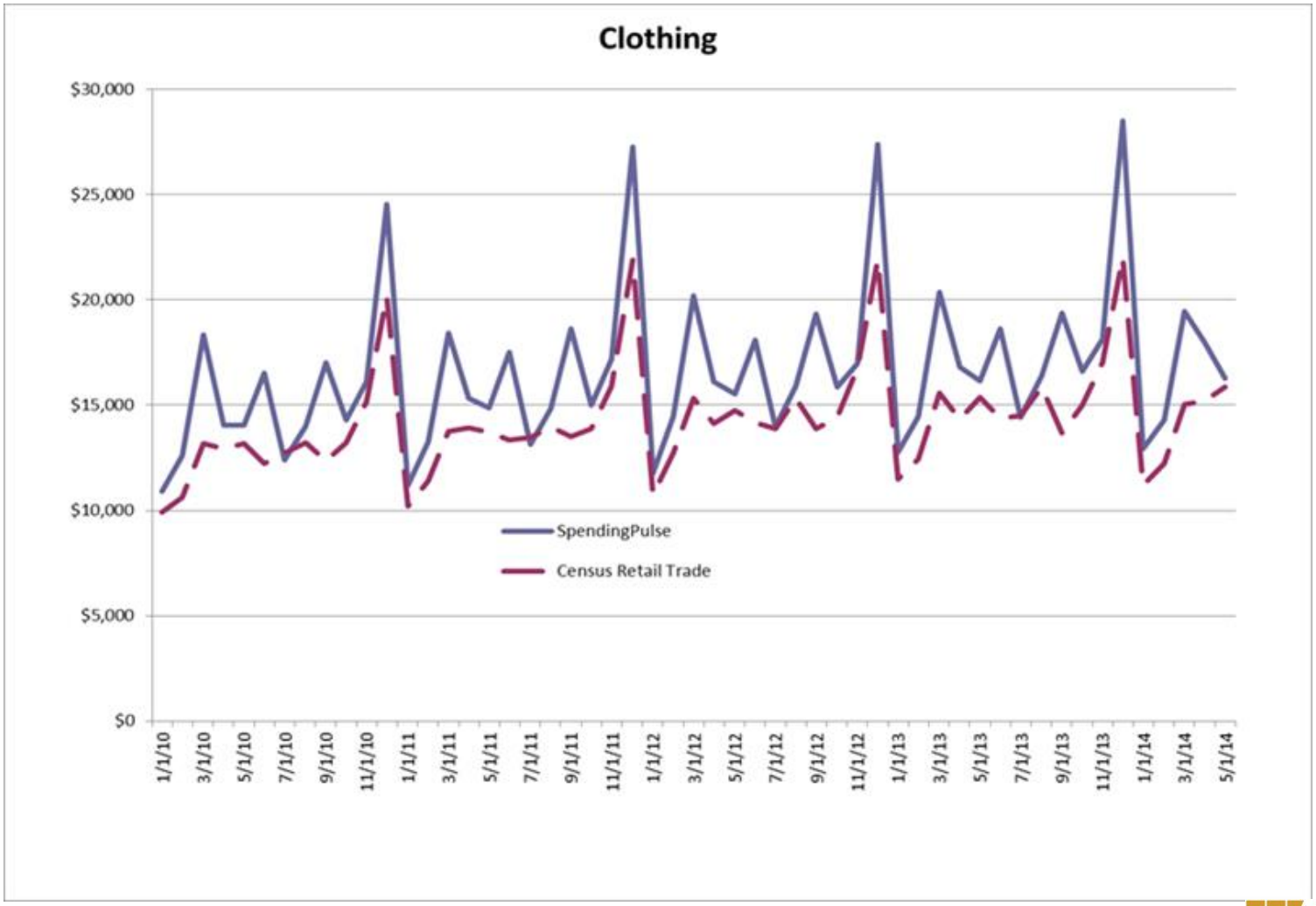
Health Care Satellite Account: Survey + Big Data



Electronic payment system data

- BEA is exploring use of credit card data to improve estimates of consumer spending and to develop estimates at the metro area and county levels
- Census Bureau initially focused on addressing data quality and mitigating the impact of declining survey response for its monthly retail sales survey
- Pilot projects: Mastercard, First Data, PayPal and Nielsen

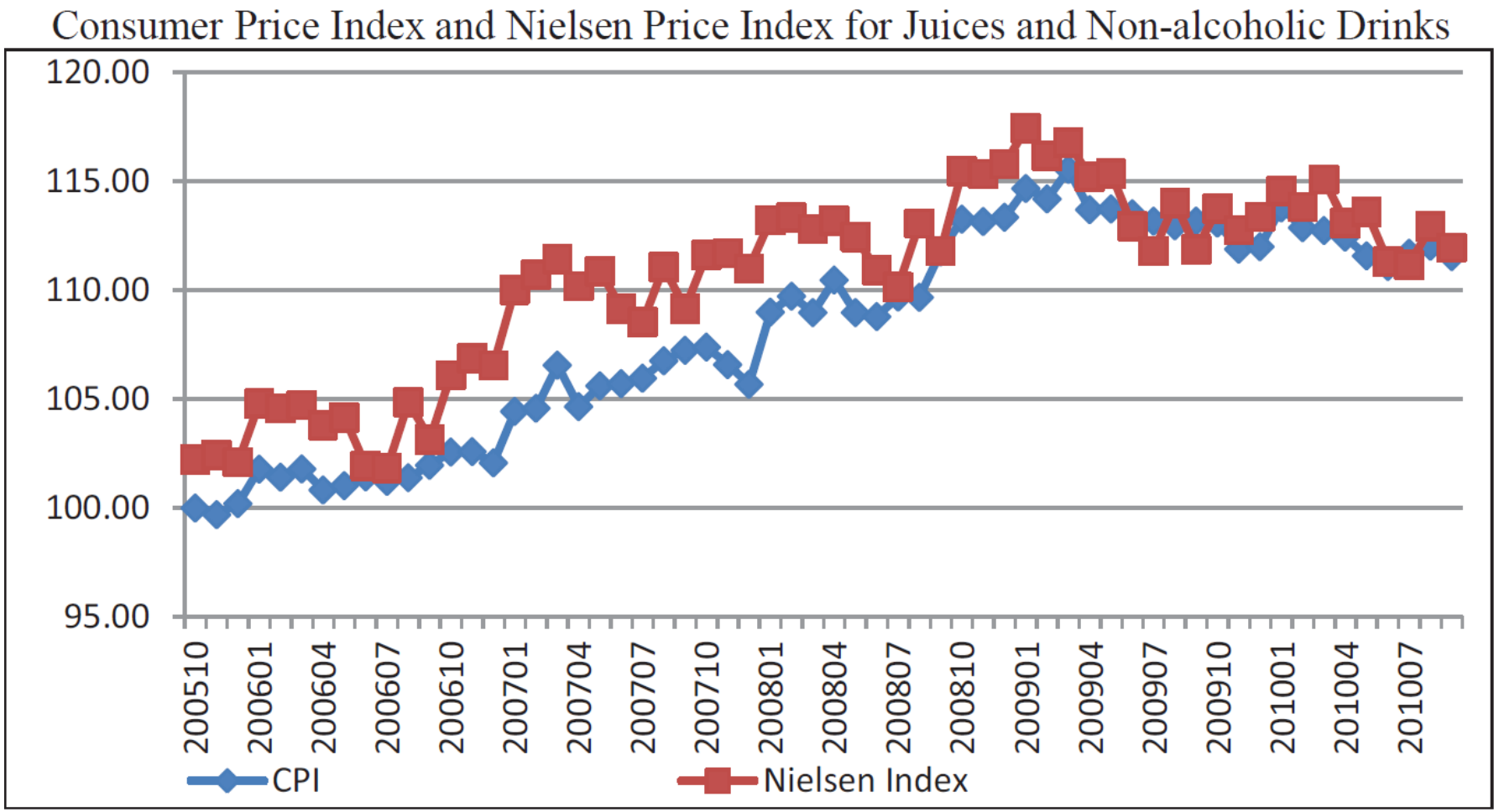
Credit Card Data for Consumer Spending



Retail Scanner Data

- Initial use of retail scanner data for BEA's estimates of consumer spending for electronic goods (TVs, audio equipment, cameras, etc.)
- BEA looking to expand the use of point-of-sale retail scanner data to estimate composition of consumer spending for type of product
- Census Bureau interested in long-run viability of producing high frequency retail statistics at detailed geographic levels
- Bureau of Labor Statistics testing web scraping and direct data feeds from retailers to measure consumer prices

Retail Scanner Data for Verification



Source: Bureau of Labor Statistics

Lessons Learned & the Future of Big Data

- All big data sources not created equal
- Use of big data requires incremental, blended approach
- Transparency issues with use in official statistics
- Need parameters to assess utility and quality of data sources (e.g. total error framework for surveys)
- Incentive structure and sustainability of big data sources
- Current and future cost considerations
- Legal, privacy and data confidentiality issues
- Public/private partnerships are key
- Need a framework with policies and procedures for using big data in official statistics