Big Data: Tackling New Projects and Exploring New Sources

Dennis Fixler

BEA Advisory Committee Meeting

13 November 2015
Background

• Using Big Data—not a big change for BEA
• Definition of big data: non-survey based data
• A broad definition would include administrative data
• Focus will be on private sources of data
• Use of administrative data has received a lot of attention: OMB directive promoting its use and setting out guidelines for its use
Background

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

<table>
<thead>
<tr>
<th>Source Data</th>
<th>Industry</th>
<th>International</th>
<th>National</th>
<th>Regional</th>
</tr>
</thead>
<tbody>
<tr>
<td>AM Best</td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Bureau Van Dijk</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Compustat</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mercer’s Health Plan</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Merger Market</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Pharma</td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>RL Polk</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Ward’s Automotive</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
</tbody>
</table>
Background

• Motivations/Goals
  – Improve early estimates
  – More regional data
  – Passive data collections
Background

Questions to ask before use

• How representative are the data?
• Do they meet the measurement objectives?
• Do they provide consistent time series?
• Are they timely and reliable?
• Are methods of compilation transparent?
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

• Use survey population weights to fold in data from different sources

  - Privately Insured → MarketScan®
  - Medicare Population → Medicare FFS 5% Sample
  - MEPS Other (e.g. Uninsured, Medicaid)
Example: Health Care Satellite Account Prices--Survey Data Only

![Graph showing price index for various health care satellite account prices from 2000 to 2012. The categories include Symptoms, Circulatory, Musculoskeletal, Respiratory, Endocrine, Nervous System, Neoplasms, and Neoplasms. The price index is normalized to 100 in 2009.]
Example: Health Care Satellite Account Prices--Survey + Big Data

![Chart showing price index for different health care satellite account symptoms over the years from 2000 to 2012. The x-axis represents the years from 2000 to 2012, and the y-axis represents the price index (2009=100). The chart includes lines for Symptoms, Circulatory, Musculoskeletal, Respiratory, Endocrine, Nervous System, and Neoplasms.]
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.
• Physician services about 25% of PCE spending on health care; health care about 16% of PCE
Exploring Athenahealth Data

• One problem related to the “convenience” nature of the data is that the size distribution of establishments leans towards small establishments.

• The chart below compares that distribution to the distribution published in county business patterns data from the Census Bureau.
Exploring Athenahealth Data

Distribution of Establishments by Establishment Size

- Share of Establishments
- Number of Employees in Establishment

- Athena Data
- County Business Patterns

BEA
BUREAU OF ECONOMIC ANALYSIS
U.S. DEPARTMENT OF COMMERCE
www.bea.gov @BEA_News
Credit Card Data for Consumer Spending

• Using credit card data collected from the mandatory survey BE-150 to inform its estimates of international travel in the Balance of Payments Accounts

• Exploring use of credit card data to improve estimates of consumer spending, and to develop estimates at the metro area and county levels
Credit Card Data for Consumer Spending

- Working with Census Bureau on acquiring and analyzing the data.
- Pilot Projects: MasterCard, First Data/Palantir
- Exploring Nielsen, and PayPal.
Monthly credit card data estimates compared to retail trade estimates

Clothing

SpendingPulse (MasterCard Advisors)

Census Retail Trade
Monthly credit card data estimates compared to retail trade estimates
Exploratory work with First Data/Palantir

Data and coverage

Aggregate Market Data
• ~50% of all U.S. Credit Card transaction spend
• Point of Sale (POS) data from 4.5MM+ 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 a subset of data
  • Five states for regional cuts
  • Anonymity and contracting restrictions
First Data/Palantir

State Share of National Sales

- CA
- IA
- MA
- NY
- TX

INDUSTRY SHARE OF TOTAL RETAIL SALES BY STATE


Metric: Industry Share
Region: State
Industry: Gasoline and Other Energy ...
First Data/Palantir

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

Metric
Growth Rate since August...

Region
State

Industry
Gasoline and Other Energy...

CA  IA  MA  NY  TX

2012  2013  2014  2015
First Data/Palantir

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

<table>
<thead>
<tr>
<th>Metric</th>
<th>Region</th>
<th>Industry</th>
</tr>
</thead>
<tbody>
<tr>
<td>Growth Rate</td>
<td>State</td>
<td>Gasoline and Other Energy Goods PCE</td>
</tr>
</tbody>
</table>

COMPARISON DATA SET

National PCE by Month
First Data/Palantir
Big Data Activities

• CNSTAT Expert meeting: Exploiting Commercial Data for Official Economic Statistics, November 19

• ICSP/FCSM: Subcommittee on Administrative Alternative and Blended Data

• UN Global Working Group on Big Data
Big Data Activities

• CNSTAT Panel: Improving Federal Statistics for Policy and Social Science Research Using Multiple Data Sources and State-of-the-Art Estimation Methods

• OECD December Workshop: Access to New Data Sources for Statistics: Business Models for Private-Public Partnerships
Summary

• Use by BEA of commercial data sources has a long history
• Current efforts geared to improving early estimates and regional data
• Many organizations involved and collaboration is crucial