Government and Big Data: What’s our future?

Erica L. Groshen
Commissioner of Labor Statistics

March 12, 2015
BLS priorities

- Data goals: **Accurate**, **Objective**, **Relevant**, **Timely** and **Accessible**

- Provide data as a pure public good
  - Like roads and clean air
  - Help all the nation’s policymakers, businesses, and families make good decisions (underpin autonomy)

- Get best we can for Nation’s data dollar
Agenda

- Government/private sector comparison
- Where are we now?
- Where to from here?
Example: BPP or CPI?

- Billion Prices Project
  - Extremely timely
  - Vast amounts of input data

- Consumer Price Index
  - Representative bundle
  - Public sampling weights
  - Known statistical properties
BPP and CPI

- Not “or”, but “and”
  - Government and private/academic initiatives are complementary, not competitive

<table>
<thead>
<tr>
<th>Government</th>
<th>Private/academic</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Transparent methods</td>
<td>• Speedy production</td>
</tr>
<tr>
<td>• Access to comprehensive and sensitive data</td>
<td>• Rapid innovation</td>
</tr>
<tr>
<td>• High survey response rates</td>
<td>• Access to some transactional data</td>
</tr>
<tr>
<td>• Objective</td>
<td>• Tailored to special needs</td>
</tr>
<tr>
<td>• Public products</td>
<td></td>
</tr>
</tbody>
</table>
How BLS already uses Big Data

- Quarterly Census of Employment and Wages
  - Sample frame
  - Firm dynamics
  - Match profit/nonprofit status

- Consumer Price Index
  - Web-scrape characteristics for hedonics
  - Transaction data

- Producer Price Index
  - Bid/ask prices for securities indexes
  - DOT baggage fees

- Survey of Occupational Injuries and Illnesses
  - Auto-code text for illnesses and injuries
BLS Big Data goals

- Enhance efficiency
- Enhance products
Opportunities and risks

- Quality
  - Bias
  - Errors, inconsistencies, poor documentation

- Access
  - Cost
  - Data stream can change or disappear

- Legal questions/constraints
  - Informed consent to link administrative data?
  - Web-scraping?
Working together

- Help us help you
  - Innovate
    - Data types and techniques
  - Support
    - Speak up for government statistics
  - Participate
    - Respond to our surveys
    - Compare findings
  - Communicate
    - What do you see, need, suggest?
How to stay in touch

- See BLS Customer Service Guide
  - Call or email our experts!
- Follow BLS on Twitter
- Sign up for email alerts
- Check out
  - The Economics Daily (TED)
  - Commissioner’s Corner blog
  - Spotlight on Statistics
  - Beyond the Numbers
Government statistics: Headlined, not sidelined

- With private sector innovation, the need for statistical agencies is greater!
  - As experts, organizers, and facilitators
  - As the “gold standard”
  - For transparency, weights and continuity

- So is the need for conversations--like this--between us
Contact Information

Erica L. Groshen
Commissioner

202-691-7800
groshen.ERICA@bls.gov
Challenges for BLS

- Measuring a dynamic economy
  - Continuity vs. change, speed vs. quality
- U.S. federal budget
  - Declining funding and certainty
- Maintaining participation in surveys
- Confidentiality and trust concerns
  - Squeezing public use files
  - Legislative restrictions on sharing
  - Public attacks on integrity
- Attracting young people to public service
Big Data challenges

- Pfeffermann’s list
  - Coverage/selection bias
  - Data accessibility
  - New legislation
  - Privacy
  - Disclosure control
  - Computer storage
  - Computation and analysis
  - Linkage of different files
  - Risk of data manipulation