Modernizing Economic Statistics: Opportunities and Challenges

NTTS2017
Brussels
March 16, 2017

Ron Jarmin
U.S. Census Bureau
Empirical social science research shifting from Surveys to Administrative Data

Use of Pre-Existing Survey Data in Publications in Leading Journals, 1980-2010

Use of Administrative Data in Publications in Leading Journals, 1980-2010

Source: Chetty (2012)
Analyses with administrative data provide facts about the economy that are not available via traditional measurement methods.

Examples include:

- Income inequality
- Intergenerational mobility
- Business, Productivity and Labor Market Dynamics
- Gig Economy
View of Gig Economy Employment: Administrative vs Survey Data

Figure 1: Household Survey and Administrative Data Self-employment Rates, 1996-2012

“Official” Economic Statistics in the U.S.: Current Status

- Structure and methods more or less the same as developed mid 20th century to serve primarily federal government purposes
- Dispersed across several agencies
- Reliant on surveys as primary means to capture source data
- Administrative records used primarily to aid surveys - e.g., sample frames, imputation, etc.
Evolving User Needs

- Desire for more timely and detailed versions of “mainline” estimates
- Desire for higher moments
- Linkage to and/or benchmarking for non-official datasets
- Filling data gaps
- More demand for microdata
  - Evidenced-Based Policy Making
  - Increasingly difficult to ensure confidentiality AND utility of PUMS files
The Incentives for Innovation in Economic Measurement

▪ Trends “pushing” agencies to innovate:
  ▪ Declining survey response rates and increasing costs
  ▪ Declining relevance of current measurement activities
    ▪ Economy evolving faster than measurement
    ▪ Increased availability of alternatives
  ▪ Flat or **declining** budgets

▪ Trends “pulling” agencies to innovate:
  ▪ New data sources
  ▪ Improved computing and analytics
  ▪ Increased demand for economic statistics
Importance of New Data Sources

- Most data that are useful for economic measurement are in the private sector.
  - But this is where the data for official designed measurements in surveys have always come from.
  - Changing technology has lead to an explosion of data.
- Challenge is to harness these new sources of data for economic measurement in a way similar to that applied to government administrative records.
“Big Data” “Source Data” Research Agenda for Modernizing Federal Economic Statistics

- Methodological
- Computational
- Policy / Legal
- User and Stakeholder Engagement
Research Not Enough!

- Self-Imposed Headwinds:
  - Architectural
    - Information Technology
    - Data Access
    - Organizational Boundaries
    - Standards (Statistical, Software, Product Characteristics)
  - Non-Architectural
    - Skills not up to date
    - Corporate Culture
    - Outdated Dissemination Practices
Modernizing Statistical Methods

- Move beyond standard survey based paradigm
- More focus on modeling -
  - Calibrated Bayes, Spatio-Temporal Modeling, etc. to combine sampled and “found” data
- Use all relevant data to build published estimates
- Machine learning especially for classification and coding (e.g., industry, product, occupation)
- Updating statistical disclosure methods (e.g., synthetic data, differential privacy) critical for modernizing data products
So, what are we doing?

- Establish Center on Big Data Research and Applications (aka Big Data Center)
- IT Architecture Review (chaired by Big Data Center Chief)
- Improving Data Access and Provision
  - Supporting the Commission on Evidence-Based Policymaking
  - Prototype IT Systems
- Several Applied Research Projects
Upskilling our Staff

- Train existing staff
- Big data class
- Text book
- Ongoing training
Upskilling our Staff

- Train existing staff
- Big data class
- Text book
- Ongoing training

Available on Amazon!!
Recruiting

- Agencies can’t compete for the best talent, so...
- Unconventional hiring
  - Insight Data Science Fellows
  - 18F
  - Presidential Innovation Fellows
    - Product and business development expertise
    - Collaboration and partnerships with academic researchers
Computational Infrastructure

- More than just new tools, new architecture and IT/Data governance needed
  - Ongoing Architectural Review
  - Progress on (in) the Cloud
  - Prototype facility at NYU – comprehensive approach to governance and provisioning of data and tools
Legislation and Policy Create Silos

- U.S. has 13 disjoint federal statistical agencies
- Data sharing between agencies often limited by law (e.g., Title 13, 26, HIPAA, FERPA)
- Data sharing law (CIPSEA 2002) only addresses sharing among three agencies and only for business data
- No clear guidance on utilizing private sector Big Data
Many of our data users are also data suppliers

- Businesses – more than households – both supply raw data to AND consume the products of official economic statistics
- View participation as a value proposition
- Many want to monetize their data assets
  - will they support public good provision?
- What can agencies do to deliver higher value products to these user/suppliers?
Back to Research

- Applied Research and Development projects demonstrate value of innovations
- Participation develops advocates throughout the agency and externally
- Research projects modernize existing programs
- Create new methods and products that can be operationalized (e.g., LEHD, BDS)
- Scaling is a major challenge due to issues discussed above
Startups’ Contribution to Metropolitan Employment
Job creation from startups as a percentage of total metropolitan employment

Startups contributed more to employment in some metro areas in the United States than others in 2014. The metro areas with the highest job creation rates from startups—3.5 percent and higher—were mostly in the South and West. Most of the metro areas in the Northeast and Midwest had job creation rates from startups that were lower than the U.S. metro average of 2.1 percent. It should be noted that while this map focuses on startups in metro areas, there is startup activity in nonmetro areas as well. In 2014, the average rate of job creation from startups in nonmetro areas was 1.9 percent.
Applied Research Activities

- Alternative source data for retail trade estimates
- Passive data collection for business surveys
- Big Data class
- Census – MIT Big Data Workshops
- NSF/Census Research Network and FSRDC research
- Cross agency (e.g., ICSP) activities
Modernizing Monthly Retail Sales Estimates

- Current estimates based on monthly survey
- Exploring how to augment estimates with non-survey data
  - Point-of-sale (scanner) data
  - Credit card transactions data
- Preliminary work is encouraging
Monthly Retail Survey and Credit Card Data
But...

- Alternative sources have issues:
  - Coverage
  - Bias
  - Expense
  - Availability

- Work so far recommends a hybrid Big Data / Survey approach
Innovation Measurement Initiative

- Collaborative research project between Census and researchers from several universities
- Integrates university administrative data on federally funded research grants with Census Bureau data assets
- Produce statistics consistent with the Census’s economic and social measurement mission and directly relevant to the data providers
IMI Background

- **Census Goals:**
  - Improve measurement of small but important sector of the economy
  - Address data gaps in the measurement of innovation and relation to economic growth
  - Learn how to collaborate with data providers to deliver data products they value
  - Prototype project that can be scaled and extended to other sectors of the economy
IMI Background

- Innovative Aspects:
  - Collaboration with the University of Michigan’s Institute on Research in Innovation and Science (IRIS)
  - Partnering with IRIS minimizes legal agreements needed at Census
  - Experiment with utilizing “fat pipe” of data for a sector of the economy
  - The university data are complementary to business and household data at Census
IRIS/IMI provides products valued by data providers

Top employee destinations by state with employee count:

1. Michigan (15,416 from 8 univs)
2. New York (13,173 from 15 univs)
3. California (11,353 from 15 univs)
4. Wisconsin (11,308 from 6 univs)
5. Indiana (8,082 from 6 univs)
6. Illinois (7,351 from 12 univs)
7. Ohio (6,774 from 10 univs)
8. Pennsylvania (5,044 from 10 univs)
9. Missouri (5,001 from 7 univs)
10. Texas (4,186 from 12 univs)
11. Iowa (3,970 from 4 univs)
12. Arizona (3,895 from 4 univs)

# of Employees staying in their home states: 65,380
IRIS/IMI shows how research is done
21st Century Economic Statistics

Survey Respondents:
- Businesses
- Households
- Organizations

Third Party Intermediaries:
- Data Brokers
- Universities
- Transaction Processors

Stat Agencies:
- Survey estimates
- Computed statistics (e.g., GDP)
- Reports / Analysis

CIPSEA, Title 13

Data Users:
- Policy Makers
- Business
- Media
- Researchers
- Public

Non Survey Data Providers:
- Federal & State Admin Data
- Other data
Final Thoughts

- Centralized Production → Federated Production
- Survey Centric Methods → Hybrid Modeled Estimates
- Siloed Governance → Corporate (or broader) Governance
- Method of Economic Measurement and Big Data Innovation at the Census Bureau focused on many “small” scale projects (aka extended hackathons)
Thank You

Ron Jarmin

Ron.S.Jarmin@census.gov