

# Modernizing Federal Economic Statistics

By WILLIAM G. BOSTIC JR, RON S. JARMIN AND BRIAN MOYER\*

\* Bostic: U.S. Census Bureau, 4600 Silver Hill Road, Washington, DC 20133 (e-mail: [William.G.Bostic.jr@census.gov](mailto:William.G.Bostic.jr@census.gov)), Jarmin: U.S. Census Bureau, 4600 Silver Hill Road, Washington, DC 20133 (e-mail: [Ron.S.Jarmin@census.gov](mailto:Ron.S.Jarmin@census.gov)), Moyer: Bureau of Economic Analysis, 1441 L Street NW, Washington, DC 20230 (e-mail: [Brian.Moyer@bea.gov](mailto:Brian.Moyer@bea.gov)). We thank John Eltinge, David Johnson, Nick Orsini and Jay Stewart for helpful comments.

Official statistical data on the structure, evolution and performance of the U.S. economy are produced by a variety federal, state and local agencies. Much of the methodology, policy frameworks and infrastructure for U.S. economic measurement evolved to meet the information needs of the federal government as it responded to the Great Depression and World War II. These measurement activities have seen expansions and improvements over the decades, but their original structure has remained largely intact.

However, there are growing concerns that the economy is evolving more rapidly than are the economic statistics we use to monitor it. The rapid pace of technological change, increasing interdependence with the global economy, and the movement away from goods to services highlight existing gaps in economic measurement and expose new ones.

At the same time, we are experiencing an explosion in the amount of information - most

of which resides within private sector organizations - that is useful for economic measurement. Before smartphones, networked sensors, Facebook and Twitter, the collection, processing and dissemination of statistical information on the economy was resource intensive and largely limited to official government statistical agencies. Unlike government statistical agencies that provide data about the economy for public benefit, private sector organizations appropriately use the data at their disposal for private benefit. The focus of our discussion here is how to leverage these rich private and public sector data sources to improve the quality of economic data available to the public.

## I. Challenges and Opportunities

Statistical agencies face four broad challenges. First, response rates for sample surveys have been falling with negative implications for data quality and costs. Second, the needs of data users are increasing in scope and complexity. They demand more timely and detailed data for characteristics of the economy not easily captured in surveys and the economy is evolving in ways that reduce

the relevance of current economic statistics and demand the development of new ones. Third, statistical agencies are facing increased competition from emerging sources of economic statistics. Finally, budgets for statistical activities are likely to remain constrained for the foreseeable future. Statistical agencies will need to find ways to do more with less. Innovation is imperative.

A combination of new data sources, improved data processing and analytics, and increased interest in economic measurement present statistical agencies with a number of opportunities to innovate and fundamentally improve and transform economic measurement. These opportunities offer the statistical agencies several strategies to address the challenges outlined above. However, more than simply shoring up the existing 20th century economic measurement infrastructure, the opportunities described below offer avenues to build an expanded system to measure the 21st century economy.

Successful innovation in federal economic statistics over the last few decades has almost always included the introduction of new source data. In some cases, this has been through new or radically improved surveys, but an even more important development has been the

increased scope and sophistication in the use of administrative records for statistical purposes.

Statistical agencies have employed administrative records for decades. A primary use was, and continues to be, to construct sampling frames. As computing has improved, we've seen increased interest in the use of administrative records for primary statistical measurement and economic research activities.

The link between trends in economic research and the growth of administrative records based statistical products has lessons for how to modernize economic statistics. Hamermesh (2013) discusses how economics has evolved into a much more empirical discipline and that we're seeing a shift from empirical work done with "ready-made" data (i.e., macro time-series and government survey data) to research using data from novel sources and from experiments. The empirical literature on business dynamics and job flows (Davis, Haltiwanger and Schuh, 1996 and Haltiwanger, Jarmin and Miranda, 2013) is an excellent model of researchers working with statistical agencies to innovate and advance economic measurement<sup>1</sup>. It shows how early interest by academic and government researchers, facilitated by support from scientific funding

<sup>1</sup> This work resulted in the introduction of the [Business Dynamics Statistics](#) series at the U.S. Census Bureau.

agencies and private foundations, can mature into full blown official statistical activities.

BEA has experimented with administrative health claims data to better measure output, spending and prices in the health care sector. As a result, BEA now produces a Health Care Satellite Account (Dunn, Rittmueller, & Whitmire, 2015).

More recently, there's been an explosion in the use of a variety of administrative datasets to study, for example, intergenerational mobility (Chetty, et. al., 2014) and Medicare (Baicker, et. al., 2014). Although none of these efforts have become regular statistical programs, organizations like J-PAL North America are promoting use of such data for program evaluations and in essence establishing special purpose statistical organizations.

Much of the recent empirical literature in economics has featured, not just re-purposed, under-utilized government administrative data, but an increasing number and variety of private sector datasets that are generated by digitized queries, transactions and sensors (i.e., the Internet of Things). The potential for creating new official economic statistics is just beginning to be explored, but as the cases described above demonstrate there are tantalizing signs of the utility of such data for economic measurement (Einav and Levin, 2014).

The measurement characteristics of these private datasets are less understood than public sector data and are subject to changes in business practices at the firms they originate from. As of now, it's still unclear how private data sources will be used for non-proprietary economic measurement activities. For instance, it may be that third party data are useful for describing changes over time, but due to coverage issues need to be supplemented with surveys in order to benchmark to aggregate totals. Nevertheless, it is clear that there's value in exploring their use. Indeed, U.S. statistical agencies are exploring possible incorporation of some of these data in experimental estimates which we discuss further below.

Statistical agencies are no longer on the technical frontier and often find themselves significantly lagging private sector organizations in data processing and computational capabilities. The techniques to process and analyze "big data" are applicable to current activities at the statistical agencies and likely central to efforts to utilize new sources of data to modernize economic measurement.

Varian (2014) describes several "big data" tools useful for manipulating and analyzing large datasets such as Hadoop and CART models. Machine learning techniques are likely

to find many applications for statistical agencies including classification, variable selection, modeling and prediction.

Private sector organizations, universities and national laboratories have made huge progress on developing the computing infrastructure to analyze big data across a number of domains. The statistical agencies lag behind, but fortunately can learn from what others have done.

Statistical agencies have traditionally controlled nearly the entire value chain for official statistical information. However, the growing amount of data useful for economic measurement that reside in private sector organizations provide agencies with new opportunities to partner with outside organizations to improve economic statistics. This is particularly true for organizations where large amounts of information from across the economy comes together as with credit card and social media companies. Below, we discuss some ways agencies are beginning to form new ways of working with external organizations to innovate and produce new economic statistics.

### **III. Incremental Strategy for Modernization**

We now turn to a discussion of how the agencies should best utilize the opportunities

emerging in new richer sources of data and powerful analytic tools to modernize, improve and expand economic measurement. We suggest an incremental approach where the agencies research and develop solutions to measurement problems in specific sectors to ensure the scope of projects remain feasible given constraints of available budget, staff skills and computing resources. We strongly suggest that agencies make fullest possible use of external collaboration to ensure project success and maximize knowledge transfer from the private and academic sectors.

Modernizing economic statistics requires the agencies to undertake a research agenda that addresses issues across four inter-related domains: methodology, computation, policy, and stakeholder engagement.

Methodological research is needed to solve how to produce timely, relevant and scientifically valid estimates of economic and social statistics from data collected from a wide variety of sources, most of which were not designed as inputs to the production of official statistics. This includes how to leverage the relative strengths of various data sources (including the agencies' own collection activities) to compensate for the weaknesses in other sources.

Research on the computational implications of the using Big Data for official statistics is

needed. The current computing infrastructure at the agencies is designed to support the needs of a survey-centric production function for economic statistics. Thus, agencies need to learn how to architect, develop and provide the hardware, software, and human capital infrastructure required to create and disseminate statistics constructed from a variety of sources including surveys, administrative sources, transaction data, social media, sensors, and so on.

Utilizing non-traditional data sources in the production of official statistics requires work to secure legal permissions and stakeholder buy-in. This includes legal agreements with data providers but also includes transparently engaging the complete set of stakeholders in the legal, privacy, and ethics space to ensure all understand the cost, benefits and risks of expanding the capabilities of the federal statistical system along these lines.

The data products produced by employing new data sources and techniques are likely to differ from traditional survey based estimates along a number of dimensions. Users will need to be satisfied that these new statistics accurately measure the phenomena we intend them to, and users may, in some cases, need to

be educated in how to properly draw inferences from estimates constructed in novel ways.

The Census Bureau's Innovation Measurement Initiative (IMI) is piloting many aspects of a modernized data collection, processing and dissemination approach for statistical agencies. Briefly, the IMI matches detailed transaction level data on research grants at universities to Census Bureau data assets at the micro level to produce new statistics on the value chain of the production and dissemination of scientific knowledge and expertise and its links to innovation, economic growth, job creation and entrepreneurship.

Some innovative aspects of this project are worth mentioning. First, the project is a collaboration of the Census Bureau and the University of Michigan's Institute for Research on Science and Innovation<sup>2</sup>(IRIS). IRIS acts as an intermediary and service provider between the universities providing data and the Census Bureau. Second, the IMI project is allowing the Census Bureau to work more directly with data providing organizations to iterate on data products that meet both organizations' needs, but still within the constraints of providing products that maintain confidentiality and are publicly available. As we envision working with variety of private sector organizations to

<sup>2</sup> <http://iris.isr.umich.edu/>

implement passive collection of broader and higher frequency data, it will be important for the statistical agencies to demonstrate value to the organizations, such as businesses, providing the data. Third, the IMI-IRIS research team is prototyping a number of innovative data products that could have application to other sectors of the economy. For example, employment and earnings outcomes of cohorts of students employed on the grants are described in Zolas et al., (2015). This work demonstrates how collecting much richer and timelier information from economic units permits the production of economic statistics of interest to key stakeholder groups, including the units themselves. Further, such statistics are simply not feasible in a survey centric economic measurement system.

The Census IMI project is but one of several initiatives now underway at BEA, BLS and Census to explore augmenting or replacing surveys with passive, automated data collection from business units. In addition to efforts to utilize third party data (e.g., credit card data) to better measure retail trade, BEA and Census have begun discussion with select companies about how to obtain much of the data (and perhaps more) currently collected on surveys via APIs or other automated means. BLS is exploring the use of similar data feeds and web scraping to collect price data. NCRN

researchers are exploring the use third party data and social media to measure consumer expenditures and employment flows (Gelman, et. al., 2015). As the scale and scope of these activities increases, it makes sense for the agencies to work together to avoid duplication and leverage unique capabilities

#### **IV Conclusion**

Given the complex issues surrounding access to, and use of, the types of data needed to improve economic measurement, the agencies will need to engage a broad spectrum of stakeholders beyond academic experts. Making greater use of private sector data for statistical purposes will mean that more and more of the data the agencies use pertaining to individuals, firms and other organizations come from intermediate sources and not from the people or firms themselves. There are a number of difficult legal and ethical issues that need to be addressed for society to trust the increasingly complex value chain of economic statistics. Thus, the agencies need to work with the legal and privacy community to consider what, if any, changes to the relevant regulatory and policy frameworks are needed.

Finally as mentioned above, the agencies will need to collaborate with a variety of organizations, but especially government administrative agencies and private sector

businesses to modernize economic statistics. This will require greater trust between the statistical agencies and these other entities so we can move past the arms-length contractual arrangements that currently dominate towards more robust collaboration. In short, the statistical agencies need to interact more fully with the economy they measure and the users they serve.

### REFERENCES

- Baicker, K., Finkelstein, A., Song, J., & Taubman, S. (2014). The impact of medicaid on labor market activity and program participation: Evidence from the Oregon health insurance experiment. *American Economic Review*, 104(5), 322–328.
- Cavallo, A. (2015). Scraped Data and Sticky Prices NBER Working Paper No. 21490.
- Chetty, R., Hendren, N., Kline, P., & Saez, E. (2014). Where is the land of Opportunity? The Geography of Intergenerational Mobility in the United States. *The Quarterly Journal of Economics*, 129(4), 1553–1623.
- Davis, S. J., Haltiwanger, J., & Schuh, S. (1996). Job Creation and Destruction. MIT Press, 87(February), 174–183
- Dunn, A., Rittmueler, L., & Whitmire, B. (2015). Introducing the New BEA Health Care Satellite Account. *Survey of Current Business*, 95(1).
- Einav, L., & Levin, J. (2014). Economics in the age of big data. *Science*, 346(6210), 1243089-1 – 1234089-6.
- Gelman, M., Kariv, S., Shapiro, M. D., Silverman, D., & Tadelis, S. (2014). Harnessing naturally occurring data to measure the response of spending to income. *Science*, 345 (6193), 212–215.
- Haltiwanger, J., Jarmin, R. S., & Miranda, J. (2013). Who Creates Jobs? Small vs. Large vs. Young. *Review of Economics and Statistics*, 95(2), 347-361.
- Hamermesh, D. S. (2013). Six Decades of Top Economics Publishing: Who and How? *Journal of Economic Literature*, 51(1), 162–172.
- Meyer, B. D., Mok, W. K. C., & Sullivan, J. X. (2015). Household Surveys in Crisis. *Journal of Economic Perspectives*, 29(4), 199–226.
- Varian, H. R. (2014). Big Data: New Tricks for Econometrics. *Journal of Economic Perspectives*, 28(2), 3-28.
- Zolas, N., Goldschlag, N., Jarmin, R., Stephan, P., Owen-Smith, J., Rosen, R. F., McFadden Allen, B., Weinberg, B. and Lane, J. I. (2015). Wrapping it up in a person: Examining employment and earnings outcomes for Ph.D. recipients. *Science*, 350(6266), 1367–1371.