

## Federal Statistics for Applied Economists

Katherine R. Smith

*JEL Classification:* Q, C81, C80

*Keywords:* Applied Analysis, Data Access, Data Priorities, Federal Statistics

The collection of trustworthy representative data and administrative information on the nation's population, economics, health, and natural resources is the ultimate public good. Thirteen Office of Management and Budget-designated Principal Statistical Agencies—including the U.S. Department of Agriculture's (USDA) National Agricultural Statistics Service (NASS) and Economic Research Service (ERS)—and over 100 other federal agencies that collect data and information to support their programs have produced more than 73,000 publicly available data sets (Executive Office of the President, 2013). Other microdata sets that are not available to the general public can be accessed by researchers working with federal agencies via agency-specific rules.

Applied food, agricultural, natural resource, and rural development economists rely extensively on federal statistics. In the most recent issue of *Applied Economic Perspectives* (vol. 95, issue 3), the analysis behind three out of eight articles employed federal statistics (Park, Costa-i-Font, and Mishra, 2013). Over half of the articles in the most recent *American Journal of Agricultural Economics* (vol. 35, issue 2) derive data from studies that rely on federal statistics (Hennessey, et al., 2013). In this article I review some major data sets likely to be used by researchers and analysts in the *Choices* audience and discuss current issues regarding federal statistics.

### Some Federal Statistical Data for Applied Economists

Perhaps the most widely used data come from the Bureau of the Census' (Census) population surveys, including the Decennial Census and the American Community Survey.

These Census products provide small area data on U.S. demographics, social, and economic characteristics of the population, and housing characteristics. Census is accommodating to those who use the summary statistics available online, and is clear about the requirements for researchers to access survey microdata (U.S. Dept. of Commerce Bureau of the Census, 2013b). ERS summarizes Decennial Census and other data that delineate "rural" or "nonmetropolitan" areas (U.S. Dept. of Agriculture Economic Research Service, 2013b).

When it comes to economic data, Census conducts a heavily used Economic Census from which it produces economic data on businesses by sector of the economy and region of the country (U.S. Department of Commerce Bureau of the Census, 2013a). The Bureau of Labor Statistics (BLS) collects wage, employment, and other economic data from employers, creates the consumer price index series, and carries out other surveys (U.S. Department of Labor Bureau of Labor Statistics, 2013). The Bureau of Economic Analysis (BEA) uses data from the Census, BLS, ERS, and others to develop economic accounts, including gross domestic product (GDP) and U.S. balance of payments. BEA also produces benchmark input-output data that are used extensively in rural development and business economic research.

NASS collects data on crops and livestock, and collaborates with ERS in implementing the annual Agricultural Resource Management Survey (ARMS).

ARMS is the only national collection of data that links agricultural practices with characteristics of the farm business, and characteristics of the household of the farm

operator. ARMS supports an array of research on farm businesses, adoption of agricultural practices, and the relationship between farm or conservation program participation and farm or operator characteristics. ERS is also the source for data series on agricultural productivity.

A variety of federal agencies collect data or have available administrative data on food, nutrition, and health. The National Center for Health Statistics (NCHS) conducts a number of surveys of health care providers and of households' health records and behaviors. NCHS' National Health and Nutrition Examination Survey (NHANES) is designed to assess health and nutritional status and is particularly unique in that it combines interviews and physical examinations of individuals. USDA's Food and Nutrition Service (FNS) makes administrative data on Supplemental Nutrition Assistance Program (SNAP) participation, participants, and implementation available, as well as administrative data on other programs it implements (U.S. Department of Agriculture Food and Nutrition Service, 2013). ERS presents "food environment" data from a variety of sources, by county, in a GIS framework (U.S. Department of Agriculture Economic Research Service, 2013a). Various units of the National Institutes of Health (e.g., National Institute on Aging) and the Department of Health and Human Services (e.g., Agency for Healthcare Research and Quality, or AHRQ) also conduct primary surveys that cover demographic, environmental, and behavioral factors.

For the study of agriculturally related natural resource, conservation, and environmental policy and phenomena, national and small area land use data are annually available from the satellite image-derived NASS Cropland Data Layer (U.S. Department of Agriculture National Agricultural Statistics Service, 2013), and

the Forest Service's Forest Inventory and Analysis (FIA) program which produces area estimates of forest land use within various subcategories (U.S. Department of Agriculture Forest Service, 2013). Some resource-linked land use data are periodically and retrospectively available from the USDA's Natural Resources Inventory, with use of anything other than gross summary data conditioned on permission from the Natural Resources Conservation Service. The U.S. Geological Survey produces water quality and water quantity inventories, and NASS has periodically conducted a Farm and Ranchland Irrigation Survey that reports water use issues. The Energy Information Administration (EIA) provides a wide range of information and data products covering energy production, stocks, demand, imports, exports, and prices, and prepares analyses and special reports on topics of current interest. This includes renewable energy uses and it is EIA that compiles data on U.S. greenhouse gas emissions. EIA data are collected and presented in a series that is fairly frequent and with great specificity as to source, geography, and industry—features that make it excellent for research on energy or the natural environment.

## Federal Statistical Issues

### Access

From the federal statistics users' perspective, a primary issue concerns the ease and degree of access to federal data. Great strides have been made in recent years to improve the casual user's search for relevant data summaries, tabulations, and cross tabulations online. All agencies provide easy access to summary data files. Many, but not all, federal statistical agencies also provide data access, search and analysis tools for use with their publically available data. Most permit the creation of customized tabulations from large data sets. The Census' American FactFinder is a tool for

identifying and obtaining summaries of data across Census surveys. The NCHS Health Indicators Interactive tool facilitates building tables that can be customized by age, gender, race/ethnicity, and geographic location to explore different trends and patterns. NASS tools permit custom tabulations. ERS provides an easy-to-use tool for creating customized cross tabulations of ARMS summary data. The Census' Public Use Microdata Sample and NCHS Statistical Export and Tabulation System give data users the tools to access and manipulate large data files on their personal computers.

### Balancing the Risks of Disclosure with Greater Researcher Access

Much original, applied economic research requires access to restricted microdata. Such data sets are restricted in order to prevent disclosure of respondents. The agencies are stewards of confidential information on individuals, businesses, and establishments and must meet the standards for protection that are required under the Confidentiality Information Protection and Statistical Efficiency Act of 2002 (CIPSEA). Researcher access to restricted data carries with it restrictions that protect confidentiality. Each agency's requirements are unique but all require, in some form or another, that applicants meet researcher eligibility requirements; a review to assure the requested statistics are appropriate for the proposed research and that the research is technically feasible; that the research furthers the mission of the host agency; that researchers be licensed and sworn in as agents of the federal government—themselves subject to penalties for disclosure; and that final products of the research be reviewed by the agency to assure inadvertent disclosure. It is not always easy to find the requirements, but searching within agencies' websites using the keyword "researcher access to microdata" seems to work. Some agencies

require that the restricted data be accessed and used only in specifically designated Research Data Centers (RDCs) where use is supervised, computers are secure, and mechanical barriers prevent casual observation of restricted data. This is true for Census, for example, which has only 11 RDCs nationwide that also host access to NCHS and AHRQ microdata. Census' policy—and similar ones for BLS, NCHS, and others—have generated complaints from researchers who must travel long distances and must obtain lodging to use the data they have been approved to access. Pity the economic researcher in, say, Laramie, Wyo., which is over 900 miles (as the crow flies) from any RDC. Other agencies, including ERS, permit remote access to restricted data within data enclaves that mimic or improve RDC security through researcher training, physical requirements for the area of access, dedication of servers, keystroke monitoring from afar, unannounced inspections, and other requirements. An alternative for remote access relies on masking data in manners that do not affect its statistical properties.

It is argued that severely restricted access prevents exploration and discovery of economic, behavioral, and other social relationships that may have high value in improving sound policy making, federal and State program implementation, and regional planning. On the other hand, there are severe consequences for federal agencies if disclosures are made. The response rates to their surveys could decline drastically as disclosure erodes public trust. At a time when legislation has been introduced to discontinue Census surveys—including the Agricultural Census and make the American Community Survey voluntary—and when suspicions of government data safekeeping are high, budget cuts or legislated survey elimination are not out of the question for any agency whose easy access policy leads to disclosure (U.S.

Congress Library of Congress, 2013a and 2013b).

### **Data Synchronization and Linking Statistics with Administrative Data**

Given the complexity of contemporary issues tackled by applied economists, the need to simultaneously use data from very different sources is not uncommon. For example, if one wanted to test a hypothesis about the relationships among crime in rural areas, SNAP participation, and a measure of housing conditions, no one agency's data will include all of the required variables. The Bureau of Justice Statistics, the Census Bureau and USDA's Food and Nutrition Service are likely candidates. But here's the rub: there has to be some mechanism for matching observations across data sets. The data would have to be consistent with one another. There are 15 federal definitions of "rural." Do all candidate data sets define rural the same way? What about consistency in timing—in what year were data collected? Are units of analysis consistent? One can rapidly become discouraged in attempts to merge data. Data synchronization would relieve some problems.

Data synchronization (also called data sharing) is the process of establishing consistency among data from different data bases and the harmonization of those databases over time. The Census, BEA, and BLS have both the authority (though CIPSEA) and a mandate to synchronize their business statistics, so that, for example, differences in coverage and industry coding for businesses surveyed by BLS and Census could be reconciled.

Data synchronization between survey data and administrative data is a special case. Administrative data are those data collected by an agency on its programs' participants for internal program purposes. They may come from different jurisdictions than the survey data (e.g., state-level SNAP administrative data), and the format

may vary across states' or other jurisdictions' record keeping. Administrative data are not originally collected for statistical purposes, and so their "fitness for use" in analysis deserves special consideration (Iwig, et al., 2013). Obtaining interagency agreement for data mergers can be a logistical barrier (if not a nightmare).

Data synchronization has multiple benefits, not the least of which is the ability to estimate unique value-added relationships. The process of linkage in itself creates new products and database infrastructure and better information is the result. However, data synchronization and linkage may also increase the chances of disclosure of confidential information and warrant close scrutiny in the security arena.

### **Statistical Program Protection and Prioritization in an Era of Budget Belt-Tightening**

As a result of budget sequestration and other budget cuts in fiscal year (FY) 2013, almost all federal statistical agencies saw their budgets decline by from 3% to 20% between FY 2012 and 2013. While all agencies have been able to maintain core programs to date, budget uncertainty and budget cuts have had several consequences. First, there have been permanent cuts or temporary suspensions of non-core surveys or programs. For example, the BLS has eliminated its programs on measuring green job products, Mass Layoff Statistics, and International Comparisons. ERS and NASS have suspended several commodity outlook-related series, and ERS has reduced the level of geographic detail for data on agricultural productivity. A second consequence is the postponement of research investments needed to cut survey costs without negative effects on response, reliability, completeness, or coverage. This has been a major concern for Census, which will have to deliver 2020 decennial census at a substantially lower real cost than

previous survey efforts would have dictated. Third, operating at lower and more uncertain budget levels has meant that agencies are either under a hiring freeze or have reduced rates of hiring, suggesting the possibility of a decline in survey data quality and timeliness as statistical and analytical staff levels decrease. Finally, agencies are restricting travel, training, and, of particular concern to researchers, research grant and cooperative research funds to prevent their core programs from being affected.

The President's proposed budget for 2014 would restore the funds lost through sequestration. Many agencies would use at least some funds at the President's 2014 level to "catch up" on postponed necessities. But there is little optimism about the 2014 proposed levels actually being realized. The conditionality of 2014 program possibilities on 2013 budget levels and the uncertainty about actual 2014 funding levels make planning very problematic for the agencies. Each is considering how to rank programs that are high priorities in anticipation of having to make some very difficult decisions in the near future. If you feel strongly about the need for a particular data series, you can tell the agency responsible for it why it is a high priority for you and what the consequences would be if it was ended or interrupted. Links to the 13 officially designated federal statistical agencies are available at The Council of Professional Associations on Federal Statistics (Council of Professional Associations on Federal Statistics, 2013).

### For More Information:

Cohen, S.H. and W. Hadden. 2004. *Issues and Impediments to Expanding Access to Confidential Statistical Agency Data: Restricted Data and Restricted Access*. Statistical Policy Working Paper No. 35, Federal Committee on Statistical Methodology Seminar. Available online at: <http://www.fcsm.gov>.

Council of Professional Associations on Federal Statistics. (2013). *Links to Federal Statistical Agencies*. Available online: [http://www.copafs.org/about/links\\_to\\_federal\\_statistical\\_agencies.aspx?](http://www.copafs.org/about/links_to_federal_statistical_agencies.aspx?)

Executive Office of the President. 2013. *Data.gov Empowering People*. Available online: <http://www.data.gov>.

Executive Office of the President, Office of Management and Budget. November 2012. *Statistical Programs of the United States Government, Fiscal Year 2013*. Available online: [http://www.whitehouse.gov/sites/default/files/omb/assets/information\\_and\\_regulatory\\_affairs/13statprog.pdf](http://www.whitehouse.gov/sites/default/files/omb/assets/information_and_regulatory_affairs/13statprog.pdf).

Hennessey, D., Taylor, J.E., Roe, B. and Khanna, M., editors (2013). *American Journal of Agricultural Economics*, vol. 35, Issue 2. Available online: [http://www.oxfordjournals.org/our\\_journals/ajae/about.html](http://www.oxfordjournals.org/our_journals/ajae/about.html)

Iwig, W., Berning, M., and Marck, P. and Prell, M. (Feb. 2013). *Data Quality Assessment Tool for Administrative Data*. Federal Committee on Statistical Methodology. Available online: [www.bls.gov/osmr/datatool.pdf](http://www.bls.gov/osmr/datatool.pdf).

National Research Council. 2005. *Expanding Access to Research Data: Reconciling Risks and Opportunities*. Panel on Data Access for Research Purposes, Committee on National Statistics, Division of Behavioral and Social Sciences and Education. Washington, D.C.: National Academies Press.

National Research Council. 2006. *Improving Business Statistics through Interagency Data Sharing: Summary of a Workshop*. Committee on National Statistics, Division of Behavioral and Social Sciences and Education. Washington, D.C.: National Academies Press.

National Research Council. 2013. *Principles and Practices for a Federal Statistical Agency, Fifth Edition*. Committee on National Statistics, Division of Behavioral and Social Sciences and Education. Washington, D.C.: National Academies Press.

Park, T. Costa-i-Font, J., and Mishra, A., editors. (2013). *Applied Economic Perspectives*. Vol. 95, Issue 3. Available online: [http://www.oxfordjournals.org/our\\_journals/aapp/editorial\\_board.html](http://www.oxfordjournals.org/our_journals/aapp/editorial_board.html).

Schlomo, Natalie. 2010. Releasing Microdata: Disclosure Risk Estimation, Data Masking, and Assessing Utility, *Journal of Privacy and Confidentiality* 2(1); pp. 73-91.

Smith, Katherine. 2013. Federal Statistics in the FY 2014 Budget, Chapter 20 in: AAAS Report XXXVIII, Research and Development FY 2014. Washington, D.C.: American Association for the Advancement of Science. Available online: <http://www.aaas.org/spp/rd/rdreport2014/>.

U.S. Congress Library of Congress. (2013a). *Thomas*. Available online on: <http://thomas.loc.gov/cgi-bin/query/z?c113:H.R.1638:>

U.S. Congress Library of Congress. (2013b). *Thomas*. Available online on: <http://thomas.loc.gov/cgi-bin/query/z?c113:H.R.1078.IH:>

U.S. Department of Agriculture Economic Research Service. (2013a). *Food Environment Atlas*. Available online: <http://www.ers.usda.gov/data-products/food-environment-atlas.aspx>.

U.S. Department of Agriculture Economic Research Service. (2013b). *Rural Economy Topic Page*. Available online: <http://www.ers.usda.gov/topics/rural-economy-population.aspx>.

U.S. Department of Agriculture Food and Nutrition Service. (2013). *Data and Statistics*. Available online: <http://www.fns.usda.gov/data-and-statistics>.

U.S. Department of Agriculture Forest Service. (2013). *Forest Inventory and Analysis National Program*. Available online: <http://www.fia.fs.fed.us/>.

U.S. Department of Agriculture National Agricultural Statistics Service. (2013). Available online: <http://nassgeodata.gmu.edu/CropScape/>.

U.S. Department of Commerce Bureau of Census. (2013a). *2007 Economic Census*. Available online: <http://www.census.gov/econ/census07/>.

U.S. Department of Commerce Bureau of Census. (2013b). *Research @ Census*. Available online: [http://www.census.gov/research/data/restricted\\_use\\_microdata.php](http://www.census.gov/research/data/restricted_use_microdata.php).

U.S. Department of Labor Bureau of Labor Statistics. (2013). *Databases, Tables & Calculators, by Subjects*. Available online: <http://www.bls.gov/data/>.

*Katherine Smith (kitty.smith@copafs.org) is Executive Director of the Council of Professional Associations on Federal Statistics (COPAFS). The Agricultural and Applied Economics Association is a member of COPAFS. The lists and discussions of agencies' data sets and topical federal data are not meant to be comprehensive. Posted comments that point Choices readers to other relevant federal data or give hints on the best way to access and employ federal data are welcomed by the author.*