

Economics Program Working Paper Series

What Role for the Private Sector in Public Statistics?

by

Robert H. McGuckin
The Conference Board

August 2004

EPWP #04 - 02



THE CONFERENCE BOARD

Economics Program

845 Third Avenue

New York, NY 10022-6679

Tel. 212-759-0900

www.conference-board.org/economics

What Role for the Private Sector in Public Statistics?

Robert H. McGuckin*

[*]

(September 7, 2004)

(Draft, please do not quote without permission of author)

1) Introduction

Federal data collection programs date from the very earliest days of our democracy. The first decennial census was conducted in 1790 with the primary purpose to apportion representatives as required by Article II of the U.S. Constitution. Federal (and state) statistical programs have grown to support ever expanding needs for information for governmental policy. For example, in 1810 questions about the quantity and value of manufactured products were added. By the mid-1800s the list of industries had expanded and information on crime, taxation, churches and other social issues were a part of the mix of questions. But it was over one hundred years before a permanent Census Office was created to ensure continuity from one census to another. (See National Academy of Sciences 2003)

While the Census Bureau remains by far the largest and most comprehensive statistical agency in terms of economic and demographic data collections, it is by no means the only federal statistical agency.¹ There are at least one hundred federal statistical programs, collecting data on workers, education, health, transportation, environmental issues, agriculture, immigration, crime, and much more. In addition, independent government entities such as the Federal Reserve Board maintain extensive data programs.

[1]

The government is not the only data collector. Many private sector organizations have extensive data collection and distribution programs. For example The Conference Board undertakes a large monthly survey of consumers with the TNS NFO survey organization and publishes various measures of consumer confidence. A similar survey, but with different methodology and measures is conducted by the University of Michigan's survey research center.

Industry and trade associations regularly poll their members and provide numerous statistics to the public domain. The Institute of Supply Management (ISM), for example, produces an important leading economic indicator, vendor performance in manufacturing, as well as a similar index for the service sector. This service sector indicator has been in place for less than a decade, but it was one of the first serious efforts to address the major data deficiencies in the service sector. Similarly, The Conference Board's index of help wanted advertising has been an important monthly indicator of labor demand for many years.

Non-profits are involved in compilation and distribution activities as well. The National Bureau of Economic Research (NBER), for example, makes available many databases created to support their analytical activities. Similarly, The Conference Board, as part of an extensive program of international productivity comparisons, develops databases in cooperation with its partner, The University of Groningen's (in the Netherlands) Growth and Data Development Center.

For-profit organizations also take part in both survey and distribution work. For the most part these activities are proprietary and center on the support of value added analytical and consulting activities. Companies like Haver Analytics and Datastream are

important distributors of data for many analysts and researchers. Many for-profit companies supply data in combination with analytical software. For example, Wharton Econometric Forecasting Associates (WEFA) distributes both data and an econometric model that is used by many forecasters and analysts.

This essay examines the role of the private sector in public statistics. The essay considers public statistics to be statistical information produced by federal statistical agencies and focuses on the role of the private sector in the production of these public statistics. It does not draw distinctions among “private” sector organizations. Moreover, particularly with regard to analytical research, the essay does not make sharp distinctions between the “private sector” and non-statistical government agencies.

This approach can surely be criticized; most would draw distinctions between for profit and non-profit private organizations, academic institutions and non-statistical federal agencies. For example, the evaluation of whether to accept an analytic project proposal to work with confidential data from a private business must take into account whether the project involves data from a competitor. This is much less of an issue for projects coming from non-profit and academic researchers funded by major scientific organizations.

As major users of public statistics, private sector and non-statistical government agencies have direct interest in the data and, perhaps more important, offer unique perspectives and advantages in subject matter research. The role of these groups in analytic work is essential, however, it is not well understood and, though substantial progress has been made, more could be done to facilitate such an understanding.

2) The Federal Statistical System

The federal statistical system encompasses a wide range of agencies and programs within agencies, and includes data about the United States from its beginnings to the present. Statistical programs and agencies generally are created to support new governmental programs and the specialized needs for data they required. For example, in 1950, at the outset of the cold war, The National Science Foundation was created to support the federal government's increasing role in Research and Development. The new foundation was given the specific mandate to:

Provide a central clearinghouse for the collection, interpretation, and analysis of data on scientific and engineering resources and to provide a source of information for policy formation by other agencies of the federal government.

The process of adding agencies and programs incrementally and linking them to subject matter areas has led to a very decentralized federal statistical system, particularly compared to statistical systems in other countries.² This approach, coupled with differences in legislative authorities, has produced a system in which legal standards, rules, and practices often differ from one agency to another. The Office of Information and Regulatory Affairs (OIRA) and the Statistical Policy Office, both in the Office of Management and Budget (OMB) provide coordination and statistical policy guidance. Also, there are a variety of interagency committees and other groups that work to standardize practices and coordinate activities across agencies.³

[2]

[3]

The differences in mandates and operating standards are especially important in the areas of data sharing and access to confidential microdata for analytic purposes, particularly where the data involved comes from more than one agency. This causes particular problems when the data involves both statistical agency records and administrative data (for example, from the Internal Revenue Services (IRS) or the Social Security Administration). Despite the differing standards among agencies and programs, a number of useful generalizations about the federal statistical system can be drawn.

First, the empirical evidence represented by the steady expansion of federal programs, suggests cogent reasons exist for a prominent and permanent public presence in data collection. Information directly required for the operation of government programs is also needed by citizens in order for them to exercise their responsibilities in a free democratic society.⁴ Moreover, since the marginal cost of supplying it to citizens is near zero, once it is collected, statistical information has many features of a public good. Further, clear economies of scale in data collection make it impractical and wasteful for more than one agency to collect the same data.⁵

[4]

[5]

Second, much of the data collected by federal statistical agencies is confidential—meaning that the identity of the person or firm providing it will not be revealed. Given this promise of confidentiality, the federal government includes questions about proprietary and private matters on its questionnaires and makes failure to respond to them punishable by law. The Census Bureau, for example, can institute legal proceedings against those who opt not to respond to questions on a number of census forms. Significantly, data collection by private organizations does not involve compulsorily

responses to questionnaires and though confidentiality is typically maintained, such promises may not withstand government requests for information. While less of a threat to survey respondents, the ability of the statistical agencies to withstand requests for data collected under a pledge of confidentiality is not absolute, especially when national security is at stake. (See, Anderson and Seltzer, 2004)

Third, statistical programs require strong analytic input from subject matter specialists in the areas they collect data. This offers one of the strongest cases for the decentralized statistical system.⁶ Decentralization puts the data collection activities close to the programs and policy makers using the information. While statisticians maybe experts in sampling designs, survey error, cognitive issues, and a wide range of practical and methodological issues involved in collecting and analyzing data, they generally are not subject matter specialists nor experts in the sources of non-sampling error associated with poorly specified concepts and imperfect measures. [6]

Fourth, there have been dramatic improvements in the availability of public statistics in the last twenty years. The computer revolution simplified obtaining and using public statistics. Further, the costs of access to published aggregate statistics and public use microdata have been reduced, and the possibilities for access to confidential microdata have expanded. The Census Bureau has a formal program, run by its Center for Economic Studies (CES) that encourages outside subject matter research proposals and other agencies have expanded their advisory and analytical programs. (See www.census.gov) Moreover, priority for such research has increased. A recent report of the Committee on National Statistics (2003), for example, devoted three of its first four

recommendations to needs for increased analytical research and access by outside researchers at the statistical agency.

3) Federal Statistical Agencies

As noted earlier, this essay considers public statistics to be those produced by a federal statistical agency, as defined by the National Research Council (2001)

A federal statistical agency is a unit of the federal government whose principle function is the compilation and analysis of data and the dissemination of the information for statistical purposes.
(Pg. 13)

Limiting public statistics to those produced by federal statistical agencies excludes from the realm of public statistics, data collected by government entities that neither make it public nor use it for law enforcement purposes. It also excludes agencies like the Federal Reserve Board and its regional offices, which produce and widely distribute statistical information (for example, monetary statistics and the index of industrial production). In other words, just because information is collected by statistical means does not mean that a government agency is a statistical agency.

The NRC (2001) designates an agency as a federal statistical agency if its principal function is the compilation and analysis of data for public dissemination. Few private sector organizations consider data compilation and public dissemination as their primary activities, nor do organizations like the Fed often have as their primary function,

the production of public statistics. The Fed's major mission involves the conduct of monetary policy, which requires substantial information. Some of the information it collects and some of it is obtained from federal statistical agencies.

It is worth noting that some organizations, both private and public, which fall outside the boundaries of the federal statistical system, produce statistics that have all the attributes of those produced by the statistical agencies. These attributes include such things as well-designed surveys, transparent operations, and data disseminated freely to the public. Despite these similarities, however, organizations like, The Conference Board make clear distinctions between their role as "statistical agencies" and their role as interpreters of economic data.

In its statistical role, The Conference Board publishes a handbook that outlines the data and sources of the series it uses in its business cycle indicators program. It includes extensive documentation of the methodology and principles used to create the coincident and leading economic indexes. It also issues a press release each month as a statistical document that provides data on the newest values of the indexes and any revisions to prior releases.

In its main role, that of an interpreter and forecaster; The Conference Board releases a separate document that outlines its analysis of the implications of the new data for the direction of the economy and provides proprietary value-added services and products, which help pay for the public information. Similar distinctions are applicable to organizations like the NBER, ISM and WEFA.

4) Statistical Purposes and Confidentiality

According to NRC (2001) statistical purposes “include description, evaluation, analysis, inference, and research.” As noted above, while statistical agencies collect data from individuals and economic units, this data may not be used for administrative, regulatory or law enforcement purposes. (See “Confidential Information Protection and Statistical Efficiency Act” 2002) This means that publicly supplied data can only be used to describe patterns and relationships involving classes of individuals or organizations. But research work often requires at least some work with the individual- and organization-specific data, both of which are generally collected with the promise to keep them confidential. Whether or not activities support statistical purposes is the key factor determining access to confidential microdata.

Confidentiality raises two substantial issues: One involves what statistics can be made public; the other involves conditions for use of the underlying confidential data for research and analysis.

- What information can be revealed without identifying individual data?
(This issue involves disclosure protection; the release of information that does not reveal confidential information on individuals or organizations.)
- What are the limits of use by those outside the statistical agency work and when is access to confidential information necessary?
(The key issue here is analytic subject matter research; that is, how and by whom is this research accomplished?)

5) Disclosure Protection

Disclosure protection -- what can and cannot be reported publicly by a statistical agency -- is rooted in both principle and practicality. Most important is the maintenance of individual and organizational confidentiality, so that the data can not be identified with the supplier. Freedom and democratic principles require protection of individual privacy, and protection of private information is essential for the efficient operation of markets. But the practical issues are also important: If confidentiality is not maintained, responses to requests for information are likely to be either ignored or the information supplied may be seriously distorted.

The disclosure issue has a long history of debate and discussion and little will be added to it here. (See Report on Statistical Disclosure Limitation Methodology (1994)) There are well-defined ways to ensure confidentiality is protected and many techniques for providing information while protecting confidentiality. The basic idea is to release as much useful information as possible without disclosing individual information.⁷ [7]

Aggregation is the principle method used to conceal individual identities for public dissemination, but public use microdata bases, noise introduction, and synthetic data are among the many methods also used to protect confidentiality. Research in this area continues. It is worth noting that the range of information that can be released without violating confidentiality is much greater for analytic results (for example, regression coefficients) than for tabular data. Indeed, special tabulations developed by analytic researchers raise troublesome disclosure issues. This means that even though new tabulations may be disclosure free by themselves, they may, when combined with

other tabulations released previously, for example as part of a regular program, violate confidential standards. Statistical agencies' divisions generally release as much tabular data as possible. Secondary disclosure analysis, which is required to alleviate such ensure against such contingencies, is especially difficult and limits the amount of additional tabulations that can be released. Use of regression coefficients and charts of estimated relationships reduces the chance of secondary disclosure and offers confidentiality protection. Release the data in the form of an estimated relationship among one or more variables exploits the error in the relationship to protect individual data points.⁸

[8]

6) Private Sector Work with Confidential Data

Statistical agencies rely on the private sector and other public sector agencies in several ways. Private sector organizations offer a broad range of goods and services to statistical agencies in the creation of survey instruments, the collection and processing of the data, and the dissemination of the survey results. Such work is part of the routine operations of most commercial statistical programs and can be purchased cheaply by public agencies, thereby either saving money or providing better quality for the same price. The 2002 Economic Census, which offered electronic filing and specialized survey instruments through a Generalized Instrument Design System (GIDS), designed and built by a private company is an example of this.⁹

[9]

The use of outsourcing pervades the private sector and it should continue in the public sector as well, as statistical agencies continue to organize themselves around a set of core competencies and leave other work for specialists. This is a natural consequence

of the information technology revolution and the specialization and division of labor that has accompanied it. Moreover, many of the technologies for handling large data collections and analysis of the data are developed by private sector entrepreneurs and it can be, as illustrated in the case of GIDS, more efficient to procure the work from outside vendors. (See McGuckin 2004)

In addition, to its role in production, the private sector is a major user of information produced by the statistical agencies. As such, private sector requirements are important factors in drawing up the statistical agency's programs, services and data products. The private sector also is involved in setting public policy agendas through lobbying activities, consulting activities, support of research and educational organizations and public reports and comments. Moreover users, such as academics, researchers working for private organizations, and businesses participate in advisory policy meetings and related mechanisms, that provide statistical agencies with much of their expertise, both subject matter and statistical.

Statistical agencies also contract with the private sector for specific analytical research projects just as they contract for other goods and services, including statistical research and advice. The practical issues in conducting such research and protecting confidentiality are very important for such work. At the Census Bureau, for example, contractors must agree to work within what is called "special sworn status," which exposes them to the possibility of severe penalty if they disclose confidential information.¹⁰ Few if any problems associated with the use of confidential information by outside contractors, consultants or researchers have surfaced which indicates that there are well-defined and practical ways to accomplish this access.¹¹

[10]

[11]

One important consideration in expanding this work beyond occasional contracts from the statistical agencies themselves is the resources necessary to support outside researchers. But historically the most important factor limiting access for analytical research has been whether the project could be accommodated under the term “statistical purposes.” In turn, this depends on how important the analytic research is to statistical agencies’ missions, why support from outside agencies is needed, and whether the private sector can be effective if it doesn’t have access to the microdata?

7) Analytical Subject Matter Research at Statistical Agencies

Statistical Agencies require support from investigators with strong analytic capabilities for a number of reasons.¹² First, analytic support provides the agency with an improved understanding of the potential problems with its data: Without working with data on analytic projects it is difficult to understand the problems with it. [12]

Second, it enables the agency, a better way to inform the public and other governmental entities, in a non-partisan way, about the economic forces that affect them.

Third, analytic work provides an understanding of the needs of customers (survey sponsors or data users), including both the executive and legislative branches of the government and the private sector.

Fourth, it often generates new data products, providing important research and development services to the statistical agency.

Finally, analytical research can lead to new surveys since analysis helps to identify areas where information is incomplete and new data is needed for resolution of an issue.¹³ Such surveys typically increase the value of the existing data by providing new and useful insights in their own right. [13]

8) Why Rely on Outside Researchers?

Since increased access to census data comes with increased risk of a confidentiality breach, it maybe logically argued that public statistical agencies would be better served by developing internal analytical programs, thus limiting the likelihood of disclosure.

There are two strong arguments against this reasoning. First, statistical agencies are not likely to develop analytical research programs on their own. Aside from difficulties in attracting the best academic and professional talent on a permanent basis, internal research programs tend to be either co-opted by the political process or starved because of inadequate funding.¹⁴ Encouraging independent outside researchers, [14] maximizes the possibility for good research thus enhancing the quality and relevance in statistical agency programs.¹⁵ [15]

Outside researchers, particularly those charged with making policy or interpreting statistical data are more likely to generate ideas and innovative ways of improving the value of the data collected, than statistical designers who focus primarily on the validity and credibility of the survey rather than its use.

Further, empirical evidence suggests that the risks of disclosure are acceptable. In thinking about what is an acceptable risk, it is important to recognize that there is always some risk of data disclosure. For example, a rogue employee may use confidential data for revenge or profit. The risk of disclosure needs to be assessed in light of the procedures in place for access and the penalties for those breaching the confidentiality rules. Taken together the procedures and penalties suggest that the risks are small.

The CES program at the Census Bureau has offered analytic researchers access to economic microdata with a carefully planned access program for some time. There have been no incidents of disclosure reported. For the most part the programs at other statistical agencies are less extensive and less formalized than at the Census Bureau. These agencies have not reported any difficulties in this area either.

Reinforcing the procedural requirements, which by raising awareness and limiting the opportunities for disclosure, whether accomplished inadvertently or by design. As discussed earlier, these include fines and incarceration as well as loss of professional standing and reduced job opportunities.

9) Analytical Subject Matter Research with Published Data

There is such a long history of research based upon published statistical data by private sector groups, academics, and government, that little comment is necessary. Such research is not just a matter of consumption by users; it is critical input to the creation of new statistics and quality improvement in series already being produced.¹⁶

[16]

Macroeconomic analysts, forecasters and business cycle researchers are major consumers of national account information from the Bureau of Economic Analysis (BEA). They also rely upon a variety of employment and productivity series produced by the Bureau of Labor Statistics (BLS). These investigators provide essential feedback concerning the quality of the statistical series produced and how well they capture the economic concepts they purport to measure. The feedback is given in many ways, from advisory committee presentations and discussions to formal papers and conferences on particular issues. The feedback helps identify gaps in available information and leads to the development of new measures.

A major factor in the shift of the economic indicators program from BEA to The Conference Board in 1995 was the need for increased analysis of the indicators. (Business Cycles Indicators Handbook 2001) While a number of the indicator series are produced by BEA, since their primary concerns and focus is on the national income and product accounts, business cycle research was a low priority. Thus, when The Conference Board assumed responsibility for producing the economic indicators it simultaneously initiated a research program to improve the indicators as a tool for business cycle analysis. Despite the front-loaded effort required to produce the indicators, the Board continues to believe that its focus on business and economic projections placed production of the economic indicators well within its general mission. The fact that the board now releases the leading indicators by about two weeks earlier than BEA did, suggests the continued wisdom of the transfer of responsibility

The same procedures – statistical imputation -- used to make the U.S. index timelier also enabled development of monthly indicators for many foreign countries. (See

Business Cycle Indicators Handbook 2001) Privatization of the foreign monthly indicators program linked it closer to the subject matter specialists and improved the program. Significantly, data used for the indicators is a mix of government and private series and are publicly available.¹⁷

[17]

Some of this input comes from independent work by the private sector to solve a problem or understand an issue. For example, The Conference Board's early work from 1918 to 1958 in the area of cost of living indexes was undertaken because member companies were interested in understanding how their salaries and wages satisfied the demands and needs of workers (See Kurian 2001) These indexes were the first continuous monthly indexes produced in the U.S. The information was important for companies during labor negotiations and for determining reasonable, across the board, increases in wages. (See Sayre 1958) Publication of the indexes was dropped in 1958 as the quality and breadth of federal programs at the BLS made the costly The Conference Board program unnecessary. However, The Conference Board maintained its interest in the indexes and convened a commission and issued a report on major recommendations for improvements in the CPI in 2001. (See McCracken, 1999)

A similar example involves the coincident and leading economic indexes. They were developed as part of an NBER program in the 1930s, shifted first to the Census in 1961 and later to the BEA in 1972. As noted above, in 1995 they were shifted back to the private sector, The Conference Board, where they could be closer to the analytic research. (See Beckman 2001) A key factor in the most recent shift was the fact that confidential microdata was not needed for the creation of the economic indexes since they rely on published aggregates from both the statistical system and the private sector.

In another example of the role of analytic research, industrial organization specialists provide much of the key analytical input to the development of detailed industry and product information collected and published as part of the economic census produced by the Census Bureau. A good deal of this work grew out of special ad hoc access programs to the Economic Census in the 1960s. For example, a major study of the extent to which firms classified in one industry produced products classified in other industries was based on special access to the Census industry data collected in the economic censuses. (See Gort 1962) One result of work with the confidential data was the creation of industry and product concentration statistics from basic data collected to support the national income accounts. These programs were discontinued in the late 1960s, for reasons that remain unclear.

These statistics were introduced in the 1963 Economic Census at the urging of many industrial organization economists concerned with the concentration of economic power in large firms and monopoly pricing arising from oligopoly. Somewhat paradoxically, the flood of research stimulated in part by this data, demonstrated that extensive industry cross-section data on business concentration and price-cost margins were not very useful in formulating public policy. The problem with these types of data is that they could not adequately distinguish between collusion and economies of scale or lower costs as explanations for the positive concentration- profits relationship observed in cross-section data. Panel data on individual firms are needed to understand this relationship. (See Jensen and McGuckin 1997)

10) The Importance of Access to Microdata for Subject Matter Research

While the importance of work with the published aggregates is considerable, without access to the underlying microdata that form the basis for the published data analytic subject matter research is severely restricted. For example, unless one looks at individual data, it is hard to assess aggregation error. Moreover, many of the most productive research projects involve longitudinal panel data. For example, the new job creation and destruction statistical program at BLS grew out of analytical work at the Census Bureau.¹⁸ The benefits of this work to statistical programs are discussed in McGuckin (1992, 1994, 1995). For more recent information, working papers are available at ([www, ces.census.gov](http://www.ces.census.gov)).

[18]

Comparison of the research of users of data from the Census Bureau's demographic programs, which carries out surveys of households and individuals, with those using data from the Bureau's economic programs area, which carries out surveys of business establishments and firms, reveals that researchers that relied on demographic data made much more progress in developing their research programs. A key difference in the way these areas developed was that microdata were available for demographic work, but not for business data. The demographic researchers had access the microdata using public use microdata files (PUMF) which the Census Bureau releases on a regular basis.

The large sample sizes involved in demographic collections make it possible to release masked microdata without fear of disclosure. In contrast, economic programs

microdata is characterized by small numbers of observations and extremely skewed size distributions at appropriate levels of aggregation and thus cannot support useful public use databases. (See McGuckin and Nguyen (1990))¹⁹ This difference is a key reason why analytic use of business data has to rely on direct access programs with confidentiality protection. [19]

This difference in availability of data influences the direction of research and data collection. The sophistication of the models, data available and insights and technical tools created to analyze them were greater in labor economics, which relies on demographic programs information than in industrial organization, which relies on establishment and business data. This despite the fact that there is no known theorem argues that the supply side of the labor market is less important than the product markets. Yet, for many years, labor market researchers had microdata on workers (the supply side of the market) available through public use files. This availability provided insights that helped fund longitudinal household surveys such as the Survey of Income and Program Participation (SIPP). There are other examples, such as the National Longitudinal Survey, conducted outside the U.S. Census Bureau at the University of Michigan. In contrast, there was little support for business unit (for example, establishment) analysis of the flows of job creation and destruction until the results of research undertaken through the Census access programs were known.²⁰ [20]

Financial economists use panels of data developed by private companies to support analysis of security markets.²¹ These panels include data on daily security prices as well as income and balance sheet information on publicly traded companies to understand the operation of security markets. They also have been used to develop new [21]

financial products (for example, options, index, mutual funds, and other “derivatives”). It is probably not a coincidence that finance, like labor economics, has been one of the more dynamic areas of social science research in the past thirty years.

Two other features of the subject matter projects should be emphasized. First, while the public use of microdata supports substantial analytical work, Census access programs are being used by increasing numbers of statistical analysts who have access to public use data. In fact, nearly one third of the projects currently underway at CES involve demographic not economic data. This demonstrates that many of the most important policy issues can only be addressed via linked data sets.²² The use of linked databases broadens the range of issues that can be dealt with and the findings typically lead to new data products and changes in surveys that improve their quality. This work also identifies gaps in existing work to be filled by new surveys and modifications to existing ones.

[22]

Second, many of the linkages that have proved most useful come from marrying administrative record sets with survey and census data. Some of the most important work has involved linkages between demographic and establishment information. It is virtually impossible to undertake such linkages without working directly with the microdata. But this work brings the differing legal standards regarding access into sharp focus.

11) Where are we?

The practical relevance of the analytic subject matter research programs involving researchers outside the statistical agencies is very strong. We now have twenty plus years of experience with a wide variety of projects and the work has brought forth many new databases and statistics as well as better understanding of the measurement issues surrounding many older programs. This supports the argument that analytical research supports statistical purposes by bringing important insights and innovation to statistical programs. Moreover, without the access programs it is not likely that statistical agencies would have been able to undertake such research.

While not all this research needs access to microdata, much of it does. Experience with access programs in the U.S. and abroad, demonstrates that the practical difficulties of confidentiality protection are solvable. Furthermore, procedures have been developed to ensure that the understanding and learning from the research is passed on to the statisticians producing the data, thus resolving a problem from the early years of the CES program.

Additional signs of progress abound. A new Research Data Center is scheduled to open in New York City before spring of 2005. It will join seven other research centers and the main center at headquarters in Washington DC. The access program at Census is expanding and projects are being approved and there seems to be more interest in subject matter research at other statistical agencies. (See www.bls.gov)

12) Looking Ahead; a Concluding Comment

Much of the easy fruit has been picked in the sense that we have identified the importance of longitudinally based economic statistics like gross job flow and entry and exit statistics. The work to date also has identified the pervasiveness differences in establishment characteristics and performance, even within narrowly defined industries. These insights will continue to guide research projects as well as the production of statistics. But we have also learned that linking data across surveys and agencies is a key to continued improvement in our understanding of the economy and the creation of new statistics. For example, developing data sets that link information on the performance of business units to the characteristics and pay of individual workers at those units are needed to evaluate training and hiring programs, business organizational issues and many other issues. Developing these kinds of data sets has raised new issues, since they typically require linkages of data across programs and agencies with differing authoritative interpretations and practices. Although resolutions of inevitable conflicts maybe difficult to obtain, they should be pursued.

Working out procedures for access has been a major issue for the Census Bureau and the IRS in the past four or five years. Completion of this process has resulted in new operational standards for access at the Center for Economic Studies. The new standards place more emphasis on the needs of the statistical agency. As such, projects must have a “predominant” purpose of increasing the utility of Title 8 data. (See <http://www.ces.census.gov/ces.php/home>)

In the past, scientific merit and an agreement by the subject matter specialist to apply the results of her findings to improve census programs, data, and products, was the operable standard. The working assumption was that an empirical subject matter project of scientific merit using the confidential microdata would provide important benefits to the data programs of the agency. In practice this standard appears to have worked to improve data programs. One can appreciate a system that values projects according to their utility to the statistical agency. However, projects with great promise for expanded understanding and only some benefits need not be excluded automatically.

While the list of criteria for deciding whether a project fits the new standard includes virtually all of the benefits that have traditionally been associated with subject matter research, the new standard still appears quite restrictive. In particular, the uncertainty of research often means the full benefits of a project cannot be foreseen at the outset of a project. In some cases it may take several complementary projects before the statistical agency receives significant benefits. Analytic projects that provide greatly expanded understanding and relatively fewer, but positive, foreseeable benefits to the data programs may ultimately yield greater benefits to data programs. Furthermore, there have been some disturbing word of mouth reports from some researchers about the difficulty of getting projects approved. Some of this undoubtedly came during the early days of the transition when some projects were dropped from the access program. Some projects appear to have suffered extensive delays and, while it is hard to document, there also appear to have been some projects that had to be abandoned.

It is clear that the process of gaining access to census data has become more standardized in recent years and requires more explicit and well-documented proposals.

This is something that was surely needed. CES was much too informal in its earliest days, in part because the resources for running the program were woefully short, particularly as the program grew and expanded to research data centers located away from headquarters and to databases beyond the economic programs area.

It is too early to draw conclusions about the ultimate effect of the new procedures. The steady flow of new projects and the pending opening of the New York Research Data Center bode well for the future. Still, it is troubling that the criteria developed to determine whether proposed projects will receive access to critical data continues to favor those employing traditional statistical methodologies rather than analytical projects that offer the promise of greatly expanded understanding.

Analytic research, including evaluation of non-sampling error and issues of whether the actual measures created from the data are conceptually appropriate for the policy issues of interest, are not explicitly mentioned. There also is no explicit recognition that output of research is uncertain. Long-term benefits and new products are often not obvious at the outset of a project. Just as business must undertake research projects that may or may not yield products or services that fit the company's business model, projects with scientific merit may eventually generate enormous benefits, even though the ex ante value to the statistical agency may be somewhat speculative.

References

Anderson, Margo, and William Seltzer. 2004. The Challenges of “Taxation, Investigation, Investigation, and Regulation.” Statistical Confidentiality and U.S. Federal Statistics, 1910-1965. Paper presented at symposium, America’s Scorecard: The Historic Role of the Census in an Ever-Changing Nation, sponsored by U.S Census Bureau and the Woodrow Wilson International Center for Scholars.

Ark, Bart van, Inklaar, Robert, and Robert H. McGuckin. 2002. “Changing Gear, Productivity, ICT and Service Industries: Europe and the United States.” Working Paper, Economics. New York, NY: The Conference Board.

Bosworth, Barry P., Jack E. Triplett. 2001. “What’s New About the Economy? IT, Economic Growth and Productivity.” *Brookings Institution* Spring (2)19-30.

Brynjolfsson, Erik, and Lorin Hitt. 2000. “Beyond Computation: Information Technology, Organizational Transformation and Business Performance.” *Journal of Economic Perspectives* 14(4)23-48.

The Bureau of Labor Statistics. April 20, 2004.(www.bls.gov).

Business Cycle Indicators Handbook, Handbook of Economic Indicators. 2001. New York, NY: The Conference Board.

Committee on National Statistics, 2001. *Principles and Practices for a Federal Statistical Agency: Second Edition*. National Research Council.

“Confidential Information Protection and Statistical Efficiency Act.” December 17, 2002. Title V, Public Law 107-347, <http://frwebgate.access.gpo.gov/cgi-bin/getdoc.cgi?dbname=107_cong_public_laws&docid=f:publ347.107>

Conly, Dalton. 2004. *The Pecking Order: Which Siblings Succeed and Why?* Forthcoming Pantheon Press. reviewed in NY Times by Emily Eakin, 2/14/2004.

Center for Economic Studies website, U.S. Bureau of the Census. March 24, 2004. <<http://www.ces.census.gov/ces.php/home> >.

Davis, Steven J., Haltiwanger, John, and Scott D. Schuh. 1996. “Gross Job Flows in U.S. Manufacturing.” MIT Press,

Factfinder for the Nation, U.S. Census Bureau. May 2000. <http://factfinder.census.gov/home/saff/main.html?_lang=en>.

“43rd Annual Report of the Conference Board.” 1959. New York, NY: National Industrial Conference Board.

- Gort, Michael. 1962. *Diversification and integration in American industry*. Princeton, NJ: Princeton university Press.
- Jensen, Bradford J. and Robert H. McGuckin. 1997. "Firm Performance and Evolution: Empirical Regularities in the U.S. Microdata." *Industrial and Corporate Change* 6(1) 25-47.
- Kurian, George T..2001. *Datapedia of the United States 1790-2005: America Year by Year, Second Edition*. Lanham, MD: Bernan Press.
- Millward, Neil. 1993. "How British Labor Economists Benefit from Establishment Surveys." unpublished working paper.
- McCracker, Paul W., James Tobin, Charles R. Hulten, Marvin Kusters, Robert D. Reischauer, Edgar R. Fiedler, Gail D. Fosler. 1999. "Measuring Prices in A Dynamic Economy: Re-Examining the CPI." New York, NY: The Conference Board.
- McGuckin, Robert H.. September 8-10, 1992. "Analytic Use of Economic Microdata: A Model for Researcher Access with Confidentiality Protection." *Proceedings, International Seminar on Statistical Confidentiality*. Dublin, Ireland. 83-97.
- McGuckin, Robert H.. March 2004. "Can Manufacturing Survive in Advanced Countries?" Executive Action, number 93. New York, NY. The Conference Board.
- McGuckin, Robert H.. June 1993. "The Importance of Establishment Data in Economic Research." Proceedings, International Conference on Establishment Surveys. Buffalo, NY.
- McGuckin, Robert H.. February 21-23, 1994. "The Creation and Use of Microdata Panels: Insights from the Center for Economic Studies' Experience." *Proceedings of the 1st Eurostat International Workshop on the Techniques of Enterprise Panels*, Luxembourg, 105-116.
- McGuckin, Robert H.. "Establishment Surveys from the End-User Perspectives," 1994 Proceedings of the Section on Survey Research Methods, American Statistical Association meetings. Toronto, Canada, pp. 146-152
- McGuckin, Robert H.. November, 1994. "Evaluating the Role of Organizational Change on Firm Performance Using Longitudinal Firm-Establishment Data." Expert testimony, OECD proceedings.
- McGuckin, Robert H.. 1995. "Establishment Microdata for Economic Research and Policy Analysis: Looking beyond the Aggregates," *Journal of Business and Economic Statistics* 13(1) 121-126.
- McGuckin, Robert H., and Sang V. Nguyen. 1990. "Public Use Microdata: Disclosure and Usefulness." *Journal of Economic and Social Measurement* 16(1)19-90.

McGuckin, Robert H., Bart Van Ark. October 2001. "Making the Most of the Information Age: Productivity and Structural Reform in the New Economy." *Perspectives on a Global Economy* Report 1301-01-RR.

Panel on Confidentiality and Data Access (U.S). 1993. "Private lives and public policies: confidentiality and accessibility to government statistics." eds. George T. Duncan, Thomas B. Jabine, and Virginia A. de Wolf. Committee on National Statistics, National Research Council. National Academy Press.

"Principles and Practices for a Federal Statistical Agency." 2nd Ed. 2003. Committee on National Statistics. National Academy Press, Washington, DC.

"Protecting Participants and Facilitating Social and Behavioral Sciences Research." 2003. National Academy of Sciences.

Sayre, Robert. 1958. "Consumers' Prices, 1914-48." New York, NY: National Industrial Conference Board.

"Statistical Programs of the United States." Fiscal Year 2004. Executive Office of the President, Office of Management.

Statistical Policy Working paper #22. May 1994. "Report on Statistical Disclosure Limitation Methodology." Statistical Policy Office, OIRA, OMB.

Triplett, Jack E. 1991. "The Federal Statistical System's Response to Emerging Data Needs." *Journal of Economic and Social Measurement*. 17(3-4)155-177.

Endnotes

* The Conference Board, robert.mcguckin@conference-board.org

¹ The Census Bureau, in addition to its own programs, undertakes data collection programs for other major governmental agencies and some private organizations. (See “Statistical Programs of the United States” (2004))

² The decentralized system is a continuing source of debate and controversy. I deal with it here tangentially.

³ There is a well codified and researched guide to principles and practices for federal statistical agencies, now in second edition, and widely used to evaluate statistical agencies and to guide the development of new ones (National Research Council (2001)). See pages 36-38 for an overview of coordination activities.

⁴ A more extensive discussion of the case for a strong presence in data collection by federal statistical agencies can be found in *Private Lives and Public Statistics* (1993). A key aspect of the argument is the idea that federal statistical agencies are independent of the policy process (to ensure professionalism and impartiality).

⁵ The cost of provision of statistical information is not free and can, in principle be undertaken through private parties as long as property rights in the information are available. Such possibilities seem difficult at best, particularly given the importance of information to democratic processes.

⁶ While programs and surveys designed to support specialized programs such as the research and development surveys created for the NSF, they also benefit from the data and analysis derived from other programs. The CES provides one method for sharing this data across programs.

⁷ Many private sector organizations conduct surveys with confidential guarantees to respondents. While many of these efforts are one-time surveys of limited scope, there are many exceptions; The Conference Board’s nationally representative sample of consumers undertaken monthly with NFO, and the many regular polls undertaken by private survey organizations for foundations, research organizations and news and media outlets, for examples.

⁸ It is virtually impossible to recover information on individual observations from a regression equation in which the number of data points is substantially larger than the number of parameters estimated. Moreover, the R-squared values are almost always not close to one. See McGuckin and Nguyen (1990), National Academy of Sciences (2003), and Statistical Policy Office (1994)

⁹ Feenstra Technologies, Germantown, Maryland.

¹⁰ The relevant law is Title 13, U.S.C., Section 214. Violations are punishable with a fine of not more than \$5,000 and imprisonment of not more than five years, or both. For mixed Census and tax data, Title 26 comes into play.

¹¹ See Census Bureau Website (www.census.gov) for an example of the procedures involved.

¹² This formulation of the benefits of analytical research is taken from my testimony to the OECD in 1994 (McGuckin (1994)). See also Triplett (1991) for a discussion of the importance of analytic research.

¹³ In fact, the U.S. Bureau attracted several surveys because the CES offered the possibility of analytic work with the survey microdata linked to the longitudinal microdata collected in the Censuses and surveys. Such linked data make it possible to design surveys in a more cost-effective manner and provide analytical benefits beyond those available in simple surveys. See McGuckin (1995).

¹⁴ In this regard, mechanisms that ensure that there is independence and wide dissemination of research results are essential. The statistical agency is not the body to deal with the quality of the research. In the absence of freedom to publish findings independently of the statistical agency, researchers are not likely to become a part of the program. See McGuckin (1992).

¹⁵ This is not to suggest that statistical agencies do not require analytic specialists. Analytic subject matter specialists are needed at the statistical agencies to evaluate projects. This ensures that the quality of the research undertaken is high and, more important; the agency is able to reap the benefits of the research projects.

¹⁶ This kind of feedback circuit is increasingly being observed in business. A substantial fraction of the productivity gains attributable to information and communication technology are associated with services sectors, particularly the intensive users of these technologies. See Ark, B, van, et.al (2002), Brynjolfsson and Hitt (2000), Bosworth and Triplett (2001), McGuckin and van Ark (1999).

¹⁷ If new series needed to be developed from the underlying microdata collected in federal agency surveys, then the work could only be accomplished if the agency did it or provided access to the data for the Conference Board.

¹⁸ See Davis, Haltiwanger and Schuh (1995)

¹⁹ There are public use data files of establishment data outside the United States. For example, Millward (1993) discusses the British Workplace Industrial Relations Surveys (WIRS) and the conditions under which they are made available. However, the United States has a different legal environment, and the WIRS data sets are based on small samples. For large samples, it is still unlikely that public use files will be released in the foreseeable future in the United States.

²⁰ Today research is just getting underway based on files linking individual data on individuals with the businesses they work with. This brings the product and labor markets together and is already yielding many new insights.

²¹ Much of this data is in the public domain because regulations require disclosure and transparency in public corporations.

²² In addition, many issues require information about the structure of the family. For example, some recent work focuses on the order and nature of siblings as a determinant of educational attainment and success. (See Conley (2004)) This is similar to the need to open the black box called the firm.