Labor force data in the next century

A BLS manager envisages possible enhancements in data collection, analysis, and dissemination, with expansion of both household and establishment surveys and much greater use of administrative data

Thomas J. Plewes

To help mark the Monthly Labor Review*'s 75th year, the editors asked both data users and data producers to speculate about programs and data needs in 2015, when the* Review *will mark its centennial. This article and the article beginning on page 9 deal with the Bureau's employment programs.*

Over the past 3 years, much attention has been focused on the shape and composition of the labor market in the year 2000. The two sets of Bureau of Labor Statistics projections to the year 2000[1] and the Hudson Institute's *Workforce 2000* report[2] have received media and academic attention far beyond the usual labor market information audience.

The long-term projections are based primarily on analysis of current labor force data. Among their many valuable functions, these projections permit us to respond to the challenges and opportunities that lie in the future. In addition, they have to be taken into account in managing the programs that produce current labor force statistics. Labor statistics, like the educational and training institutions they serve, must be fine-tuned to assure that adequate measures are in place as the work force of the future evolves. It is sometimes said that statistical programs face a special challenge. To be useful, they must stay ahead of the trends, for their function is to identify events and measure those trends as they happen. Staying ahead, in turn, means that the programs must be in place before the projected changes they measure occur. The challenge to all statisticians who deal with projections—but to BLS statisticians in particular, given the Bureau's reputation for providing reliable, useful statistics on a timely basis—is to pay close and constant attention to the projections. While the projections are based on the Bureau's best current data, it is recognized that those data are themselves only as good as past projections and resource investments have allowed them to be

Timing of change

Statistical programs require long lead times before fundamental changes in approach and scope can be implemented. Hence, the likeliest scenario for the turn of the century is that most labor statistics programs will appear to their users much as they do today. Change will be evolutionary, rather than revolutionary.

The long horizon of change in major statistical efforts is exemplified in the ongoing effort to modernize the Current Population Survey (CPS), the premier household survey in the labor force field. A joint BLS–Census Bureau committee has been meeting for more than 2 years to plan for a post-1990 redesign of the survey. The opportunity to redesign this household survey comes just once each decade, because the information from the decennial census is needed for the redesign. The planning and budget process in the Federal Government is such that formal

Thomas J. Plewes is Associate Commissioner, Office of Employment and Unemployment Statistics, Bureau of Labor Statistics.

plans for work that must take place this year (fiscal year 1990) in preparation for the post-1990 census redesign had to be included in agency budget requests formulated back in April 1988. Even though preparation begins in earnest in 1990, the incorporation of decennial census results into the sample redesign will delay the completion of the redesign until 1995.[3] From inception to implementation, the modernization of the CPS will require at least 6 years. The next opportunity to introduce a significantly revised survey will be in the year 2005, with planning required to be completed as early as 1998.

We will see the emergence of a paperless environment that will bring about vast changes in data collection, transmission, editing, and publication.

Components of change

The scope and depth of changes in the labor statistics programs will depend on the future course of the three principal components of any statistical program: methodology, technology, and resources. In each of these, the exact direction of change is uncertain. Accordingly, what follows is a vision of the future; whether reality will bear it out remains, of course, to be seen.

From today's vantage point, a scenario of quantum advancement in methodology and technology is quite probable, while the limiting factor will clearly be resources—both human and financial.

Methodology. Methodological enhancements are cascading into the Bureau's statistical programs as never before, driven by concerns over quality, made possible by advancements in computing technology, and sustained by innovative statistical design practices. The pioneering work beginning to come from the Bureau's Cognitive Laboratory is an example. In this laboratory, an interdisciplinary team is testing the cognitive aspects of questionnaires on individuals and business respondents, challenging and refining not only the questions, but also the ways in which they are asked and the concepts to which they pertain.

Model-based estimation techniques, introduced into the computation of State employment and unemployment estimates just this past year, hold promise for future applications. More changes employing these techniques will come as we learn to incorporate theoretical concepts that found their first practical use in engineering applications to the large, complex statistical operations that produce labor force information.

Innovations in methodology will change the way that the Bureau develops and tests questionnaires, draws survey panels, computes estimates, and measures reliability.[4] The torrent of methodological innovations that pours out of the minds of theoreticians practically daily is fully expected to increase. At a minimum, innovations in methodology will be needed to maintain a sound statistical base for the work of the Bureau in a resource-constrained environment.

Technology. As with methodology, the pace of technology is fast increasing, and its impact is only now becoming understood.[5] Within the next decade, we will see the emergence of a paperless environment that will bring about vast changes in data collection, transmission, editing, and publication. At the same time, advances in artificial intelligence and expert systems will change forever the way statistical agencies code, edit, and analyze data. These technologies, for the most part, are available today. The issue is not where the programs will go with technology, but how far and how fast it will carry them.

The Bureau's relationship with its reporters will be rethought. Because even the smallest companies will computerize their work force data, there will be mounting interest in direct links between companies and the Bureau's data bases. Direct links to reporters will be only one of a number of radical differences in the Bureau's methods of data collection.

One of the most promising techniques for cutting down on the total amount of labor in data collection and for reducing errors is computer-assisted telephone interviewing. In a number of Bureau surveys, interviewers call respondents, asking questions that appear on computer screens and entering the interviewees' responses into the computer. An extension of this facility is touch-tone data entry, which more than 1,000 companies are currently using monthly to transmit data to BLS cooperating State agencies. Talking to the computer is the next step. Though still in the early stages of testing and development, with about 100 live cases under investigation, the technology for voice collection—in which a machine-generated voice can ask questions, record responses, and convert voice answers into machine-readable text—is right around the corner.

Distribution of data will be enhanced by the evolution of standard data exchange formats. The capabilities of optical disks and their stand-alone successors will permit wide distribution of high-volume historical data. Even if there is no expansion of the ability of statistical agencies to electronically disseminate their data directly to users, all users will benefit from the increasing capability of BLS to produce master data bases in formats that are widely useful.

Resources. Enhancements in methods and technology certainly promise to make opera-

tions more efficient, and to substitute capital for some of the more labor-intensive statistical activities. However, the achievement of these breakthroughs requires up-front investment, which may not be possible in a period of tight budget constraints and competing needs.

A sharper focus for demographic data

The projections to the year 2000 paint a portrait of a labor force emerging over the next decade that is quite different from the labor force of the past. This new labor force will be increasingly composed of minorities, women, and mature workers. In other words, the groups that will grow the most are those which customarily have experienced the greatest difficulty in the labor market, have suffered more labor market-related economic hardship, and presently are most difficult to measure using current concepts, techniques, and procedures.

Despite the emergence of complementary surveys, such as the Census Bureau's Survey of Income and Program Participation, the monthly Current Population Survey, conducted by the same Census Bureau for BLS, will continue to be the primary analytical vehicle for understanding these groups and their labor force trends. Like other household surveys, the CPS best illuminates information on the demographic composition of the work force, the interaction between the family and the work force, and the reasons for the behavior of the work force. More than other surveys, the CPS provides the size, scope, flexibility, and continuity needed to depict changes in the status of groups of workers.

To the outside observer in the year 2015, the basics of the CPS will probably appear to have stayed much the same. It will still be a monthly survey using rotation panels in a 4–8–4 configuration and pertaining to a reference week that includes the 12th of the month. Yet, the survey in the year 2015 will be substantially different from the survey of 1989. This difference will be the result of the earlier mentioned comprehensive redesign and modernization that will follow the 1990 Decennial Census of Population and Housing. Ideas already under consideration for this major redesign of the survey include the following:

- Modernizing the questionnaire to sharpen measurements and improve information on occupations and industries of workers.
- Redesigning the sample to incorporate updated materials from the decennial census, thus increasing the efficiency of the survey.
- Enhancing the reliability of State estimates by expanding the sample in mid-decade to permit sample-based estimates for all States on a monthly basis.
- Improving coverage of minorities.
- Automating data collection by widespread use of computer-assisted collection technology.
- Extending the ability of the survey to focus longitudinally on the labor force by improving the basic coding of information from one snapshot view of the labor force to another.

Plans are also under way to initiate a separate longitudinal panel to follow persons continually over long periods of time, perhaps 24 months. The data on aggregate changes in labor force status from one month to the next (gross flows), which are now compiled but not generally used in analysis, will be improved to allow for analysis of the causes of movements in the data. This improvement should aid in understanding the sometimes erratic movement in the over-the-month employment situation.[6]

To further assist in identifying the underlying economic importance of month-to-month movements in the employment and unemployment statistics, the process of adjusting for seasonality of the data will be enhanced. The Bureau is testing the concept of concurrent seasonal adjustment—that is, revising seasonal adjustment factors each month as new data become available—and appending the test series to the Commissioner's monthly testimony before the Joint Economic Committee.

The new processing environment also will change the way in which CPS data are used. Building on current trends, the Bureau will make CPS microdata, which are now available in public-use format (that is, purged of personal identifiers), more immediately available to all users on an ongoing basis. This enhancement in data access is expected to permit users to generate as much or as little detail as is desired. Freed from the limitations imposed by restricted data access, a flood of innovative analyses by university and private-sector economists and statisticians will supplement the Federal Government's analytical efforts.

Advances in artificial intelligence and expert systems will change forever the way statistical agencies code, edit, and analyze data.

Even with this more elaborate household survey data system, the scope of the surveys and their coverage of the work force will continue to be severely limited. Household surveys are expensive to initiate and maintain; they impose a considerable burden on their respondents and require cadres of skilled enumerators, who are expected to be increasingly difficult to hire and retain. Thus, traditional surveys like the CPS have a constrained potential for further expansion.

Less frequent, but more complete, enumerations are also difficult to justify. For example, a

mid-decade census, required by law for some time, would fill the void in geographic detail between decennial census years. However, given the current budget climate, such a census is not very likely. As a result, for the foreseeable future, labor force analysts will increasingly be forced to look to administrative data, together with supplementary surveys of limited scope and quick turnaround, to extend work force information.

Extended use of administrative data

A powerful new longitudinal data base on employers will be available that will support research into the process of job creation.

Despite the great promise of extended use of administrative data for labor market analysis, the United States has thus far seen mostly limited direct use of administrative records in the labor field. This stands in sharp contrast to practices in the Scandinavian countries and much of Europe, where extensive use has been made of an elaborate set of administrative records for functions that range from estimation of unemployment to time-use studies and analysis of income distribution. In large measure, the lesser use of administrative data in this country stems from a perceived tension between the administrative necessities of Government operations and the statistical opportunities the data represent. Statistical uses have been deemed an expensive nuisance by managers of programs that were established to pay benefits, collect taxes, or perform other governmental functions.

Building on a carefully developed and firm foundation laid only over the past few years, that perception is changing in the work force information field. Increasingly, the fit between the information gathered by the States as they manage the unemployment insurance system and the needs of State and Federal agencies for statistical data has come to be seen as symbiotic. The unemployment insurance data system (known as the ES-202 system, after the name of the form used in aggregating the data) has emerged over the past few years as a powerful means of understanding the workings of the labor market, identifying the process of job creation, focusing on how businesses operate, and detailing the labor force characteristics of persons. Improvements in the quality of information about establishments and persons served by the States' programs has strengthened both the statistics and program management. Thus, while many purely Federal sources of administrative data have lost much of their utility, because of both restrictions on their use for nonadministrative purposes and lack of investment in their statistical infrastructure, the Federal-State unemployment insurance system holds the promise of future potential.

Over the past several years, BLS and cooperating State agencies have devoted considerable resources to expansion of the unemployment insurance data base. These efforts began in the 1970's with standardizing claims information and improving the identification of the place of residence of the unemployed for purposes of improving local area unemployment statistics. Later, the process of affixing current industrial classification coding to business records was enhanced when all States adopted a standard 3-year update cycle and converted to a BLS-developed verification format for collecting the desired information. More recently, the emphasis has shifted to obtaining workplace information on all establishments, thus improving the information on where business activity actually takes place.

The potential for an ever more significant role for these administrative data in the labor market information portfolio is significant. Early in the next decade, a powerful new longitudinal data base on employers will be available that will support research into the process of job creation. Analysts will be able to trace the birth, expansion, structural transformation, and demise of American businesses as never before, and survey designers will be able to incorporate this information into more efficient sample designs. The new power thereby gleaned to study the life cycle of businesses using a broadly based list with quarterly coverage of all industries and sizes of establishments will transform the way that analysts look at U.S. industry. The potential of this data base for economic development and other applications at the Federal, State, and local levels is unbounded.

A number of steps will have been taken to bring this transformation about, under the umbrella of the Bureau's Business Establishment List program. The ultimate objective of this program will be to provide the basis for a single standard listing of nonagricultural businesses for common use by Federal Government statistical agencies. It is expected that the Bureau will be able to begin sharing the enhanced data in 1992.[7]

Technical and procedural changes will foster additional uses of the data. The processing environment will shift from a focus on maintaining aggregates to maintaining microdata for the establishments. Tape transmittals of input data from the States and thence to users will be a thing of the past, with new transmission media allowing for better access to, and archiving of, the data. These initiatives will also mean that data will be available at least 2 months earlier than today's capabilities allow.

The classification system will, of course, have to be modernized to match the potential of

the data system. Standard Industrial Classification (SIC) coding will be standardized and simplified. In this regard, the 1997 SIC revision may provide a new way to look at industrial coding that could include both multiproduct and conventional means at a greater level of detail (5–7 digits), thus making it possible to better associate product with place.

High technology in business surveys

Although improvements in the administrative data base will increase the usefulness and speed the inception of this census-type information, the natural lag in administrative records processing and the limitation on the kind of data available in the administrative files suggest the need for a continued monthly survey of businesses to measure employment, hours, earnings, and other characteristics of economic interest. The survey program that now serves that role—the Current Employment Statistics program—is a massive Federal-State operation involving mail and telephone collection of data from nearly 350,000 businesses each month, generation of two preliminary estimates followed by a "final" estimate, and an annual benchmark revision that reanchors the survey to the administrative data base. Steps have already been taken to prepare this survey to reflect technological advancements, including computer-assisted telephone interviewing, touch-tone data entry, voice recognition, computer-assisted personal interviewing for response analysis surveys, and many more useful enhancements as they come on line after testing and demonstration.

Data collection from business establishments will be even more closely tied to the unemployment insurance administrative data base and share most concepts with it, particularly in the earnings area. In the future, the survey could provide current data for all metropolitan statistical areas, expanding beyond the 180 areas for which data are now available monthly. The 2 months of preliminary estimates now required because of shortfalls in survey receipt when estimates are prepared will be reduced and perhaps even eliminated as technology and improved sample design allow BLS and cooperating States to focus on getting responses from key firms in a timely manner.

The survey operation will be extended to include more kinds of compensation and working conditions of employees. Currently, the survey collects only hourly earnings information, yet BLS studies have shown that nonwage items are an increasingly important part of workers' compensation.[8] A model for collection of these additional items exists in the Statistics Canada program; a similar program could be implemented on an annual basis following a period of testing and evaluation.

Occupational data with meaning

Although the pace of labor force and employment growth over the next decade is expected to slow, an analysis of the changing occupational structure of employment leads to the inevitable conclusion that the combination of industry employment trends, technological change, and other factors will increase our need to enlarge our understanding of occupations, including their skill requirements and demographic profiles. Today, the major sources of information about the occupational profile of the work force and of the Nation's employers are, respectively, the Current Population Survey and the Occupational Employment Survey. Many of the general improvements to the CPS—for example, better coding schemes and the ability to compare information from previous interviews with the current collection of interviews—will enhance the occupational data collected from individuals, and updated Census Bureau occupational classifications will increase the value of these data. But on the other hand, occupational data collected from households may be expected to continue to have notorious shortfalls. These data suffer from improper specification, skill level inflation, underreporting, and imprecision. Accordingly, for many purposes, occupational analysis, particularly in an industrial context, will continue to rely primarily on information collected from establishments.

Analysts will be able to trace the birth, expansion, structural transformation, and demise of American businesses as never before.

The Occupational Employment Survey of the next century will be a more generalized survey, serving additional users and permitting collection of additional data elements, such as earnings and demographic characteristics. The survey will be more precise, based on a completely new Standard Occupational Classification system that will have been developed for the Government in the early 1990's by a group under the leadership of the Office of Management and Budget. The Dictionary of Occupational Titles program, managed by the Department of Labor's Employment and Training Administration, will be closely integrated with this classification system. The Dictionary provides a basis for reconciling worker skills and traits, educational requirements, and occupational identification. With the vast changes in the skills that will be required for the jobs of the future, an upgraded Dictionary is a necessary first step toward making the worlds of education and work coincide more closely.

A depth of local detail

The likelihood of a continuation of areas of plenty and of poverty within the United States, within the States, and within local jurisdictions makes high-quality collection, analysis, and dissemination of State and local data paramount. In cooperation with the State agencies, the Bureau now produces unemployment statistics for some 5,000 subnational areas on a monthly basis. It is difficult to contemplate extending that detail further without a significant improvement in the availability of raw data about small-area labor markets. Therefore, any improvements achieved will be in quality and timeliness.

Current local area unemployment estimates are based on a "handbook" procedure, which assigns values to the local area based on those of larger areas for which data are available. The Local Area Unemployment Statistics program will be modernized, with estimates for all areas based on procedures akin to the model-based estimation that has been used to compute estimates at the State level for the past 2 years.

The local estimates will be processed in a multilevel environment using the next generation of PC-based software the Bureau is providing to its State partners in the State Systems Project. This environment will enhance the quality of the data and allow data production of State and local estimates to be speeded up significantly.

IN SUMMARY, the labor force scenario for the year 2015 and beyond envisions three core sources of data with analytical and dissemination programs built around them in a satellite configuration:

- A monthly household survey.
- An unemployment insurance–covered wage and employment data base at the establishment level, designed to yield quarterly aggregations of the universe of business establishments and a current monthly survey of establishments, to obtain employment linked to the unemployment insurance universe together with hours and earnings; an annual survey of those same establishments to obtain additional compensation items; a periodic occupational employment survey in some detail; and an annual occupational employment survey to collect demographic data.
- An unemployment insurance individual-record data base that will provide the basis for local unemployment estimates and information on layoffs, plant closings, and other actions of public interest.

The labor force data programs of the future will be much more oriented toward individual records, utilizing the emerging power of the computer to process those records on demand to provide tailored aggregations of characteristics of interest. This new power will simplify data storage and the job of the analyst, but will require artificial intelligence or other advanced applications to reduce the nearly infinite number of possible cross-tabulations to a manageable, understandable few. The microdata orientation will also place a new pressure on the Bureau and other statistical agencies to develop new means of making public-use files available while protecting the confidentiality of the data.

Finally, the new environment in the year 2015 will have caused us, in the interim, to reflect together on the fitness of our decentralized national and Federal-State cooperative statistical systems to meet the challenges posed. As we thus reflect on the strengths and weaknesses of our administrative system in the light of a long-term vision of the data for which we are stewards, we stand assured that we can, and will, provide for the coming expansion in the availability of high-quality labor force data. □

Footnotes

[1] The most recent BLS projections to the year 2000 are found in a series of five articles in the November 1989 *Monthly Labor Review*.

[2] William B. Johnston and Arnold E. Packer, *Workforce 2000, Work and Workers for the Twenty-first Century* (Indianapolis, The Hudson Institute, 1987).

[3] William P. Butz and Thomas J. Plewes, "A Current Population Survey for the 21st Century," *Proceedings, Fifth Annual Research Conference* (Washington, Bureau of the Census, 1989), pp. 12–13.

[4] George Werking, Alan Tupek, and Richard Clayton, "CATI and Touchtone Self-Response Applications for Establishment Surveys," *Journal of Official Statistics*, vol. 4, no. 4, 1988, pp. 349–63.

[5] Carl J. Lowe, "Statistical Processing in the Year 2015—What Can We Expect?" paper presented at Fourth International Roundtable on Business Survey Frames, Newport, United Kingdom, November 1989.

[6] Butz and Plewes, "A Current Population Survey," p. 11.

[7] Brian MacDonald, "Progress Report: Bureau of Labor Statistics," paper presented at Fourth International Roundtable on Business Survey Frames, Newport, United Kingdom, November 1989.

[8] *Employment Cost Indexes and Levels, 1975–1989*, Bulletin 2339 (Bureau of Labor Statistics, October 1989).