A data user's look back from 2015

An academic analyst of BLS employment programs steps 25 years into the future, where the kinds and amounts of data and the technology for assessing them are greatly expanded

Daniel S. Hamermesh

To mark the 75th year of the Monthly Labor Review, *the editors invited several producers and users of BLS data to speculate on changes they foresee in the next 25 years. The author of this article looks back from an imaginary vantage point in the year 2015.*

For someone like me, whose academic career began in 1965, empirical research on the labor market in 1989 was phenomenally easy. But by today's (2015) standards, our 1989 methods were primitive technology. Today's young labor economists surely are as incapable of appreciating the difficulties, in the 1980's, of conducting empirical research using data tapes that had to be obtained and manipulated with great effort as their young counterparts in 1989 must have been of appreciating the difficulties of doing research on data that, in the 1960's, had to be hand-copied and keypunched onto small cards. No doubt they would be equally flabbergasted by the paucity of data available in 1989. The dual revolution—in the technology of using data and in the kind and amount of data available—has, to some extent, resulted from decisions made in BLS during the 1990's.

Perhaps the most important of these decisions was the recognition that problems of confidentiality of establishment data could be overcome and those data made readily available to researchers outside of Government. This obviated the need for such worthy, but partial, approaches as the Census Bureau's Longitudinal Employment Database, an annual panel of manufacturing establishments that was only accessible to researchers who became sworn Government employees and who worked with the data at the Census Bureau. The change was facilitated by the development, in the late 1990's, of essentially error-free transmission mechanisms from BLS computers to individual users around the country via fiber-optic methods. As a result, researchers now can sit by their home or office computer and operate on data files located at BLS, extracting the data they desire or performing statistical analyses on BLS data files. BLS computers are programmed to prevent the export of data or the calculation of summary statistics that might violate promises of confidentiality. Blanket prohibitions on access are no longer needed—restriction is on a case-by-case basis, making access interactive and immediate.

We academic researchers pride ourselves on working on timeless issues; but very little research in the social sciences is timeless (or even very long-lasting). In the 1970's and 1980's, the difficulties of obtaining data made it necessary for us to do much of the testing of theories about the labor market on data that were 10 or even 20

Daniel S. Hamermesh, professor of economics at Michigan State University and research associate, National Bureau of Economic Research, expects to be retired from regular teaching and research in 2015.

years old. Technological developments have changed that and added a new currency to academic research. The large sets of microeconomic data that we now can obtain usually include information that is no more than 6 months old. Since 1996, Current Population Survey enumerators have been able to code data into their portable computers during their interviews for transmission immediately after; and establishments participating in employer-based surveys have responded electronically since 1993. The only lag in the process is the brief time needed to ensure the data are error-free before BLS allows public access.

By 2015, the monthly Current Employment Statistics survey should be enhanced to obtain information by occupation and by sex.

Establishment survey expanded

Naturally, the increasing ease of access to nearly-contemporaneous data stirred users' interest in the kinds of data that were accessible, and no more so than with the previously neglected establishment data. The scope of these data was greatly expanded, although the data collection was mostly an extension and rationalization of what already existed in 1989. The monthly Current Employment Statistics survey (BLS-790), which provided the published series on weekly earnings, hours, and establishment-based employment by industry, was enhanced to obtain information by occupation and by sex. With some encouragement and guidance, employers have been willing to submit the required data.

Most important, information on employee benefits, training and other nonwage labor costs, on the output of each establishment, and on job vacancies was included in the data. The neglect of nonwage costs had made the BLS-790 data increasingly irrelevant for research on the determinants of compensation and for studies of employers' decisions about hiring and firing. With the collection of information on these costs, researchers in 2015 are able to conduct serious studies of how these costs affect employers' decisions about adjusting their work forces. With the collection of data on output by establishment, we now can explicitly link shocks to product demand to the costs of changing employment. The creation of a continuing sample of job vacancy information has provided a long-needed analog to the unemployment information in the Current Population Survey and has fulfilled the promise of the aborted job-vacancy programs of the late 1960's. With all these changes, immediate information on the structure of employment in relation to wages and product demand thus became available to researchers.

In the 1980's, economists realized that employment change was largely a reflection of plant openings and closings.[1] With access to these establishment-based data, researchers have been able to identify the dynamics of employment, *both in continuing plants and in those that opened or closed,* in a systematic and comprehensive way. Instead of merely charting the sizes of flows of jobs, the inclusion of employment cost and output data has enabled us to measure the determinants of these flows as well. We can now study worker displacement at the appropriate level, that of the individual plant.

The stock-in-trade of labor economists has always been the analysis of wage differentials. The development of these accessible, large-scale sets of establishment data has enabled us to study wage differentials, or, more correctly, compensation differentials, at the plant level and to include the characteristics of the individual establishments. This has allowed us to test theories of macroeconomic adjustment based on so-called efficiency wages that were in vogue in the 1980's and 1990's. It has permitted us to dispose of a variety of questions on the hoary issue of compensating wage differentials (and, given the nature of research, to create new questions that cannot yet be answered by existing data).

The excessive concentration of data on the manufacturing sector has finally ended. With fewer than 15 percent of jobs in the U.S. economy remaining in manufacturing in 2015, this broadening of information has been vital to understanding trends in wages and wage differentials. It has provided the chance to study questions about the determinants of wage differentials, flows of job opportunities and their causes, and employment adjustment outside the narrow context of manufacturing. In short, these broader data have enabled researchers to develop and test theories of labor demand generally, not just within manufacturing.

Household, establishment data linked

As crucial as these developments have been in redressing the imbalances between household- and establishment-based data, they would not by themselves have been revolutionary. What has been revolutionary is their link to household data. By allowing checks on individuals' reports of their earnings and hours, this link has enabled BLS to develop programs that removed much of the substantial measurement error that caused some of us to question research on wage determination using the household-based data of the 1970's and 1980's.[2]

Even more important than the enhanced quality of the household data has been the ability this

linkage has given labor economists to study wage and employment determination *in a market* rather than merely from the employer's or worker's side. With information on samples of particular firms' employees, we can examine how changes in the demand for labor lead to adjustments in behavior of the worker and members of the worker's household. We finally have been able to identify the relations that generate wage differentials, so that we can actually specify changes in a worker's or employer's behavior and predict their impact on the structure of wages. The development of data during the 1970's and 1980's enabled us to produce sophisticated tests of complex theories of labor supply and demand that had their origins between the 1930's and 1960's; the development of linked establishment-household data in the 1990's enabled us to test the contracting theories that had their origins in the 1970's and 1980's.

The resources, both monetary and time, devoted to expanding establishment surveys and making them accessible were not, of course, free. But there was sufficient political will to spend resources in an area that loomed increasingly important. At the same time, enough resources still were available to finance the continued expansion and refinement of the Current Population Survey. Expansion of the CPS enabled researchers to examine the detailed structure of labor force dynamics in a few of the largest labor markets. Its increased size allowed researchers to test hypotheses about the changing structure of the labor force and of unemployment within particular demographic groups which the smaller sample sizes had previously not permitted. The enlarged sample even allowed us to trace the impact on local labor markets of large-scale plant closings, so that the CPS data could function like the European job registration data in this regard. With a larger sample, we also had sufficient observations to make the linkage with the establishment data noted above.

Longitudinal CPS

Some of the longitudinal household data sets financed by the Federal Government, but organized and collected privately, continue to this day; but much of the academic interest in them has been supplanted by attention paid to the Longitudinal CPS that started in the mid-1990's when outgoing rotation groups began to be interviewed systematically. Initially, these interviews were only for 2 years. That soon expanded; some of the households now have been in the Longitudinal CPS for 10 years. The obvious advantages of a larger sample size engendered many new possibilities for studying subgroups of the labor force. These data have enabled researchers to examine the determinants of transitions between labor force states with a precision that was impossible using the earlier longitudinal household data sets. No longer do labor economists debate the purely "counting" questions of the relative importance of incidence and duration of unemployment over the cycle, or of the magnitude of transition probabilities by demographic group among employment, unemployment, and nonparticipation. These data have changed the focus of the debate to allow us to construct and test economic theories of the determinants of these probabilities and of unemployment incidence and duration within particular demographic groups.

The Current Population Survey has continued to function as a vehicle to which supplements that provide data to examine topics of current interest can be attached. If anything, these have increased in number, as researchers and civil servants have recognized the ease of obtaining data in this way. (The effect the supplements have on the quality of the regular CPS data is not yet certain.) As examples, one supplement responded to concerns about nonmonetary aspects of employment by asking detailed questions about working conditions, employees' perceptions of the nature of their work, and their knowledge of conditions in the business. This gave policy analysts the information necessary to examine the incidence of various safety and health problems that the old establishment-based data could not disclose. It gave academics the ability to formulate models based on workers' perceptions and expectations, thus allowing us to examine much better the intermediating role of expectations in economic behavior. A supplement in 2005 concentrated on workers between ages 45 and 55 and obtained information about the economic status and demographic characteristics of the "baby-boom" generation in middle age.

We academics still pay too little attention to, and are still woefully ungrateful for, the quality of the data provided to us by government agencies. Particularly noteworthy were the improvements in the quality of data that have occurred since 1990. Partly these have resulted from the improved technology of data handling—the substitution of direct data entry for most of the paperwork has reduced error rates considerably. Partly, too, these have occurred because of the increased sophistication of the algorithms for assigning values to missing data points. Perhaps most important, though, they have been generated by the increased concern of all data users that the raw material of their analyses be as free of error as is possible, and the recognition of

In 2015, expansion of the CPS will enable researchers to examine the detailed structure of labor force dynamics in a few of the largest labor markets.

policymakers that the research that occasionally informs their endeavors should not be based on unnecessarily dirty data.

The burgeoning supply of data has improved research in yet another way: less professional payoff is acquired by those who obtain data and perform a few simple analyses, and more is now given to those who think and analyze carefully. Admittedly, this was not true in the 1990's when the new wealth of data resources was a novelty; but as the novelty wore off, and the increased ease of doing research became apparent, ideas, not just manipulation of data, became more heavily rewarded. This paralleled the earlier moves away from the novel, but uninformed, "regression-running" of the 1960's, and from the excessive concern with the structure of error terms by "laborometricians" in the 1980's.

Looking ahead, 2015–40

The one constant among economists is our desire for more data. New expectations and hopes spring up as soon as our old requests are satisfied. Just as the development of economic theory stimulated and was stimulated by the creation of new sets of data before 1990, and between 1990 and 2015, no doubt that synergy will affect the development of data during the next 25 years. As before 1990, and as between 1990 and 2015, emerging social issues will focus researchers' attention on generating new economic approaches to thinking about them, and will create a demand for new types of data. I doubt that the next generation of economists will be any more satisfied with the data at their disposal than we are with our data, or our predecessors were with theirs. Just as we are far more fortunate in this regard than economists working in the late 1980's, though, I have no doubt that our successors will look back at us and marvel at the underdeveloped state of our analyses and the data that underlie them. □

Footnotes

[1] For example, see Timothy Dunne, Mark Roberts, and Larry Samuelson, "Plant Turnover and Gross Employment Flows in the U.S. Manufacturing Sector," *Journal of Labor Economics*, January 1989, pp. 48–71.

[2] Greg J. Duncan and Daniel Hill, "An Investigation of the Extent and Consequences of Measurement Error in Labor-Economic Survey Data," *Journal of Labor Economics*, October 1985, pp. 508–32.