

This PDF is a selection from an out-of-print volume from the National Bureau of Economic Research

Volume Title: Labor Statistics Measurement Issues

Volume Author/Editor: John Haltiwanger, Marilyn E. Manser and Robert Topel, editors

Volume Publisher: University of Chicago Press

Volume ISBN: 0-226-31458-8

Volume URL: <http://www.nber.org/books/halt98-1>

Publication Date: January 1998

Chapter Title: Existing Labor Market Data: Current and Potential Research Uses

Chapter Author: Marilyn E. Manser

Chapter URL: <http://www.nber.org/chapters/c8356>

Chapter pages in book: (p. 9 - 50)

1 Existing Labor Market Data: Current and Potential Research Uses

Marilyn E. Manser

Major new research questions and policy issues concerning labor markets have arisen in recent years. An overriding set of issues involves the perception that many jobs have become less secure and newly created jobs may not be “good jobs” on a number of dimensions. Analysis of these issues requires understanding of contingent work and other nontraditional work arrangements, the pattern of individual job changes and career growth, the process of job destruction and creation, the structure of compensation, and policies within firms. At the same time that there appear to have been major changes in the labor market, there has been little change in the concepts used, the types of information collected, the surveys employed, or the way data are processed and made available.

The purpose of this paper is to set the stage for addressing the current needs for labor market information. The first part provides background for the conference by describing existing government data on employment, unemployment, and compensation and the major purposes of these data and by analyzing their uses in recent labor economics research. The second part of the paper is interpretive and more forward looking: it examines other possible uses of the data and problems in responding to changing needs for data.

Marilyn E. Manser is assistant commissioner for employment research and program development at the Bureau of Labor Statistics, U.S. Department of Labor.

The views expressed are those of the author and do not reflect the policies of the Bureau of Labor Statistics (BLS) or views of other BLS staff members. The author is grateful to Charlie Brown, Dan Hamermesh, and a number of BLS staff members for helpful comments and discussions on a previous draft; to Frank Stafford for detailed discussions about his earlier review of data use in labor economics research; and to Karen Kosanovich for expert research assistance on the analysis of the recent economics literature. Any errors are the sole responsibility of the author.

1.1 Existing Data Sets and Usage

The “official” U.S. data on labor markets are produced by the Bureau of Labor Statistics (BLS). BLS data are also used for labor economics research and analysis, along with various data sets produced by other agencies. To provide some historical perspective, table 1.1 lists the BLS’s present labor market data programs, in chronological order of introduction.

The BLS predates the Department of Labor, having been established in 1886 after receiving considerable support from the labor reform movement.¹ For instance, Samuel Gompers of the American Federation of Labor testified in Senate hearings during 1883 that a national bureau “would give our legislators an opportunity to know, not from mere conjecture, but actually, the condition of our industries, our production, and our consumption, and what could be done by law to improve both” (quoted in Goldberg and Moye 1985, 3). The initial focus of the bureau was on prices and, in the labor area, on wages for various occupations and industries. What may seem in more recent years like an over-emphasis on certain types of workers (production) and on certain industries (manufacturing) probably has its origins in historical concerns about “working class” men and women in what were earlier emerging industries.²

Information on the workforce was obtained for the first time by the Census Bureau in decennial censuses of population conducted in the 1800s. Data were obtained based on the “gainful worker” concept, which counted persons aged 10 and over based on their usual occupation; there was no unemployment measure. During 1914–15, the BLS considered and conducted surveys on unemployment in part of New York City. These were not continued, but a new program to measure employment and payrolls in certain establishments began publication in late 1915 with five manufacturing industries. This was the beginning of the Current Employment Statistics survey program. Just after World War I, interest in unemployment heated up again, but no major changes in data collection concerning unemployment occurred for a number of years afterward. Recommendations for expansion of the employment data to cover non-manufacturing industries, to collect information on hours worked and some characteristics of workers, and to improve statistical aspects of the survey were made as early as 1926.³ Data collection for some nonmanufacturing industries began in 1928. In response to policy concerns about unemployment and research with sample surveys that attempted to classify the population as employed, unemployed, or out of the workforce, a new conceptual framework was

1. See Goldberg and Moye (1985, 2–3). This section draws heavily on information provided in Goldberg and Moye.

2. A focus on production workers and manufacturing has existed for some other countries as well—e.g., Great Britain. Additional study would be needed to see if this is typical.

3. A committee was formed by President Hoover to study the factors underlying employment and practical measures that could be undertaken to reduce unemployment. The result was a study by the National Bureau of Economic Research, the Russell Sage Foundation, and the American Statistical Association that was published in 1926.

Table 1.1 BLS Ongoing Labor Surveys

First Data Collection	Originating Agency	Program	Date Transferred to BLS
1800s	BLS	Occupational pay surveys	
1915	BLS	Current Employment Statistics (CES or 790)	
1935	Department of Labor—Manpower Administration	Unemployment insurance data (ES-202—Universe File)	1972
1940	Work Projects Administration transferred to Bureau of the Census in 1942	Current Population Survey (CPS)	1959
1948	BLS	Collective bargaining	
1966	Department of Labor—Employment and Training Administration	National Longitudinal Surveys of Labor Market Experience (NLS)	1986
1971	BLS	Occupational Employment Statistics (OES) Survey	
1976	BLS	Employment Cost Index (ECI)	
1980	BLS	Employee Benefits Survey (EBS); earlier surveys of benefits were also done	
1981	BLS	Hours at Work Survey	

adopted for the 1940 census.⁴ Also in 1940, a new survey—the predecessor of the Current Population Survey—was begun with labor force classification based on an “activity” concept, which included only those persons who were working or seeking work.⁵

Thus, over 50 years ago, a framework of data on employment, unemployment, and wages had been established based on concerns of policymakers, unions, and academics. Numerous specific changes occurred over the years, leading to the configuration of labor market data programs and other useful research data sets that exists today. In order to help inform the discussion about current needs, the next sections provide an overview of these data sources.⁶

1.1.1 BLS Labor Surveys

This section provides a brief discussion of BLS labor data programs.⁷ It does not attempt to fully describe all content of a survey or to assess the important

4. Professional economists inside and outside government and organized labor both played important roles in the call for and development of the employment and unemployment series at various points from 1915 through the 1950s.

5. This paragraph is based in part on the National Commission on Employment and Unemployment Statistics (1979, 21–23).

6. The reader may also be interested in Antos’s (1983) discussion of concepts for analyzing labor costs and his survey of BLS information relating to labor costs that existed in the 1970s and early 1980s.

7. The Occupational Safety and Health program is excluded from this survey.

issues of survey methodology and data quality. Documentation on all of the BLS series, with the exception of the “special surveys,” is provided in Department of Labor (1992).

Household Surveys

Current Population Survey (CPS). Empirical labor market studies probably rely most heavily on the CPS, a monthly survey of about 50,000 households collected by the Census Bureau for the BLS. This survey is extremely timely: information on labor force status—employment, unemployment, out of the labor force—for the week of the 12th of the month is collected the week of the 19th. The official monthly unemployment rate is calculated from this information and is released about three weeks after the reference period, usually on the first Friday of the following month. A key federal economic indicator, this measure is among the measures most watched by the policy and business communities. Tabulations of labor force status for various demographic groups are also regularly provided based on CPS data. Although establishment surveys are understood to yield better quality measures of industry and occupation, the CPS is the primary source for research on employment or wages by occupation or industry whenever demographic information is needed as well.

A major revision of the CPS, begun in 1986 and conducted jointly by the BLS and the Census Bureau, was recently completed; the official estimates have been based on the new survey since January 1994. In addition to updating the sample design based on the 1990 Census of Population and improving collection methodology and processing, this revision improved question wording. There were also conceptual changes in certain measures (discouraged workers, part time) in accord with recommendations of the National Commission on Employment and Unemployment Statistics (1979), often called the Levitan Commission. This revision completed implementation of the commission’s recommendations concerning the monthly CPS except for those concerned with longitudinal data (see below).

Many research and analytic studies based on the CPS use data from the March income supplement funded primarily by the Census Bureau, which provides detailed annual income information and recall measures of employment and unemployment over the past year. Various special supplements, including the Displaced Workers Supplement, have also provided important labor market information.

National Longitudinal Surveys of Labor Market Experience (NLS). The NLS provides data over a long period of time for the same individuals for four “original” cohorts begun in the mid- to late 1960s, namely, older men, mature women, young men, and young women, and for the youth cohort (now called NLSY79) begun in 1979. Data collection for the older men and young men was discontinued in the early 1980s, but the older men were reinterviewed in 1990 with funding from the National Institute on Aging. Data for the ongoing original cohorts and the NLSY79 are collected and disseminated under a con-

tract with Ohio State University. Currently, interviews with about 3,100 respondents from the mature women's cohort and 3,400 respondents from the young women's cohort are collected in a combined survey conducted biennially. About 9,000 respondents are still interviewed in each round of the NLSY79, which was conducted annually through 1994 but is now biennial.

The NLS contains a rich body of labor market and related information and permits the analysis of issues that cannot be explored with cross-sectional or short-term longitudinal surveys. The NLSY79 data include comprehensive work history information, measures from the Armed Services Vocational Aptitude Battery, and information on an especially wide variety of special topics. The National Institute for Child Health and Human Development has sponsored collection of a unique body of fertility-related information and a set of assessment instruments administered every other year beginning in 1986 to the children of the female respondents. These assessments provide information on the child's development and the home environment. The Department of Labor's Employment and Training Administration, and now the BLS, have made available NLSY79 data with identifiers and important variables for local areas through special agreements with research institutions.

Among changes in questionnaire content in recent years, the NLSY79 section on employer-provided training was reinstated and expanded. Questions on special topics have also been included; for example, questions relating to internal labor markets, focusing on promotions and job ladders, were asked in 1988–90, and questions on contingent work were included in the 1994 survey. A regular BLS publication with descriptive analyses, *Work and Family*, was initiated to bring important results to the attention of persons beyond the NLS's traditional users in the academic research community.

A fiscal year 1995 budget expansion provided funding for a new youth cohort, NLSY97, with data collection beginning early in 1997 under a contract with the National Opinion Research Center (NORC). Data will be collected annually for a sample of about 10,000 youths aged 12–16 as of 1 January 1997.

Establishment Surveys and Administrative Records Data

Employment and wages covered by unemployment insurance (ES-202 program). The ES-202 program maintains data collected from administrative records covering all employers subject to the unemployment insurance (UI) and unemployment compensation for federal employees (UCFE) programs. This effort is a cooperative endeavor of the BLS and the state employment security agencies. Using quarterly data received about five months after the close of the quarter, the BLS summarizes data on employment and wages and salaries for workers covered by UI and UCFE laws, who account for 98 percent of wage and salary civilian employment.⁸ Annual average employment and "wages" (a broad measure of cash earnings per employee) are published for

8. Estimates are available initially about four months after receipt of the data. Some minor revisions are made to the data.

virtually all four-digit SICs at the national and state levels. Also, monthly employment and quarterly wage data are now available at the national, state, and county levels. The ES-202 does not provide hours data. It is used as a benchmark for the Current Employment Statistics (see below). In addition, microdata from the ES-202 program are used as the sampling frame for all BLS establishment surveys except the 790.

The program provides information to the Employment and Training Administration and the various state employment security agencies for administering the employment security program. It is used by the Bureau of Economic Analysis as the basis for estimating a large part of the wage and salary component of the national accounts. An effort is under way to develop a longitudinal establishment-level file. There has also been discussion of a proposal to create a wage records database.

Current Employment Statistics (CES or 790) survey. The CES is an establishment survey conducted monthly in cooperation with the state employment security agencies. It provides very timely data for the pay period including the 12th of the month on employment for all employees and women workers and, separately, for production workers in goods-producing industries and nonsupervisory workers in service-producing industries.⁹ Preliminary monthly estimates are released along with the CPS estimates, usually on the first Friday of the following month. Data on payrolls and hours paid are collected for production and nonsupervisory workers only (about 81 percent of workers). Information on overtime hours is collected for manufacturing industries. Data are collected from about 390,000 reporting units, permitting publication of current data for the United States as a whole, for all states, and for 275 metropolitan areas, with extensive industry detail.¹⁰

Together with unemployment data from the CPS, the employment data from the CES are generally the first major economic series to be released each month. The employment estimate is used in the Index of Coincident Economic Indicators, and the manufacturing average weekly hours series is used in the Index of Leading Economic Indicators. The CES data are important to business and policy users who are concerned about current industry conditions because they are the only very timely information available at the detailed in-

9. Although the CES predates the statistical use of the ES-202 data, one can now think of the CES as a survey that "moves" the ES-202 employment data, i.e., provides preliminary estimates that will later be updated by ES-202 data, provides monthly rather than quarterly payroll data, and provides hours information.

10. The CES sample is unique among BLS surveys in not being probability based. Formerly, this was not seen as a problem by the program, given the focus on aggregate estimates, for two reasons. First, the sample size has been extremely large; it was about doubled during the late 1970s and mid-1980s, and now the reporting units include 40 percent of all persons on nonfarm payrolls. Second, the employment data are benchmarked to the ES-202 (although the payroll and hours estimates are not benchmarked). The American Statistical Association Panel (1994) recommended that the BLS implement probability sampling for this survey, and the BLS is currently conducting research in this area.

dustry level. In addition, the CES earnings data are used in preliminary estimates of the national accounts and the hours data are used in estimates of productivity.

The CES payroll and hours data are used to construct an average hourly earnings measure. This measure will be affected by changes in employment of less experienced workers, in occupations, and in differences between full-time and part-time status (to the extent that such workers are paid at different rates) that occur over the cycle or over time, so it may not proxy well movements in the cost of hiring labor; see the description below of the Employment Cost Index, which is designed for this purpose. (For some important purposes, such as analyzing wages received by the typical worker, use of an index that measures the cost of the effective quantity of labor would not be appropriate.) Comparisons of average hourly earnings between goods-producing and service-producing industries are problematic because the concept of workers included differs between these sectors. Finally, benefits are excluded. For all of these reasons, the CES wage data are probably less widely used by researchers and business economists than the employment information, as noted in American Statistical Association Panel (1994, 55–59).¹¹

Occupational Employment Statistics (OES) survey. The OES was begun in 1971 in cooperation with the Employment and Training Administration and 15 state employment security agencies to obtain national, state, and area occupational employment estimates by industry. It presently surveys about 725,000 establishments on a three-year cycle. Recently, 15 state employment security agencies began collecting occupational wage distributions. The OES has been used primarily for educational planning. Beginning with the 1996–97 collection cycle, plans are for the OES to expand to survey each industry every year and also to collect wage information.

Compensation surveys. The Employment Cost Index (ECI) was developed in response to the need for a broad measure of the change in wage costs. First published in 1976 for wages and salaries and in 1980 for total employee compensation, it measures the rate of change in employee compensation, which includes wages, salaries, and employers' costs for employee benefits. The ECI is published quarterly, one month after the reference date. The sample comprises about 4,200 establishments in the private sector and 1,000 establishments in state and local government.

As a fixed-weight, Laspeyres index of the change in the price of labor, the ECI provides an important picture of compensation changes that eliminates the effects of employment shifts among industries and occupations. That is, it provides a measure of the change in the "price" or "cost" to the employer of a

11. The BLS is currently conducting research on possibilities for also collecting a more comprehensive earnings measure in the CES.

unit of labor. Thus the ECI series is preferable to an average hourly earnings-type measure for many analytic purposes. However, industry and occupational detail are not extensive. The ECI is more comprehensive in terms of benefit costs than are other U.S. series. Since 1987, annual estimates of compensation cost levels have been published; this series is constructed using the ECI cost change data and current weights for industry employment from the 790 and for occupations from the ECI. In addition to its use by economists and policy-making agencies, the ECI is used as an escalator of labor costs. Under the Federal Employees Pay Comparability Act of 1990, one phase of the federal pay-setting process now calls for a general pay increase for all white-collar federal employees that is tied to changes in the ECI.¹²

The present annual Employee Benefits Survey (EBS) began in 1980, in response to the Office of Personnel Management's desire for data on benefits. Expansion of coverage has taken place since then. Presently, data on the incidence and characteristics of employee benefits are collected in odd-numbered years for medium-size and large private establishments and in even-numbered years for small private establishments and state and local governmental units. The sample includes 4,400 establishments in the private sector and 1,300 governmental units. EBS data are used in analyses by various government agencies and business organizations, and in labor negotiations.

Over the years, data on wages for selected occupations and industries have been collected in a variety of occupational pay surveys. The effort was restructured in 1990, with a focus on data collection for Office of Personnel Management use in federal pay setting, where in addition to a general pay change there is a provision for adjustment in pay by area. The Occupational Compensation Survey Program (OCSP) presently collects information on wages in 46 selected occupations, by work level, from about 18,000 establishments in about 160 localities. This survey replaced predecessor surveys—namely, the White Collar Pay Survey, whose predecessor was the Professional, Administrative, Technical, and Clerical survey; the Area Wage Surveys, which provided information for selected occupations in major standard metropolitan statistical areas;¹³ and the Industry Wage Surveys, which formerly were conducted in various industries on a three- to five-year cycle. The OCSP data are compatible with the Area Wage Survey data.

It is interesting to note that the primary use of the Industry Wage Surveys and Area Wage Surveys in recent decades was not the same as for most economic statistics, that is, by policymakers and the business community in general, nor was it for research purposes. Instead, the primary users were compensation specialists in firms or unions, who used them for "wage determination purposes." Interestingly, Rees (1993) argues that in practice wage comparisons

12. In practice, since 1978 an alternative plan for federal pay has always been substituted for the payment specified by the process using wage data.

13. In addition, special, more limited, area-type wage surveys have been and still are being conducted for, and used by, the Department of Labor's Employment Standards Administration in administering the Service Contract Act of 1965.

are made in both nonunion and union situations and if no suitable survey exists firms will conduct one.¹⁴

The BLS has begun an effort, called "Comp 2000," which is considering possibilities for integrating the ECI, EBS, and OCSP programs.

Hours at Work Survey. Begun in 1981, this survey of about 5,500 establishments collects quarterly and annual ratios of hours at work to hours paid. It is used to adjust the hours paid by industry information from the CES for use in BLS productivity measures.

Negotiated wage and benefit changes. This program, which was eliminated in 1996 in response to budget cuts, prepared information on current changes in wages and benefits agreed to in collective bargaining. Provisions of agreements were published in *Compensation and Working Conditions*. Monthly listings gave the company, employer association, or governmental unit in which a change had occurred, the union involved, the general wage increase and effective date, and related information on the contract. The BLS prepared series on wage rate changes in major collective bargaining settlements (those covering 1,000 workers or more) and on compensation changes in the largest settlements (those covering 5,000 workers or more in all industries and 1,000 workers or more in construction).

Special surveys. Beginning in the late 1980s, some special establishment surveys and pilot studies have been conducted by the BLS. The first were surveys of contracting-out practices in selected industries. Special surveys of day care and drug testing have also been collected. A pilot survey that collected information on demographic characteristics of a sample of employees within establishments, conducted as part of the White-Collar Pay Survey, is of interest for research purposes (see Bronars and Famulari, chap. 13 in this volume). A pilot survey on job vacancies was also conducted, but a full survey, which would have been quite costly, was not funded.¹⁵ Most recently, two establishment Sur-

14. Rees uses this in arguing for the role of "fairness" in wage determination, but other things could also be responsible, i.e., imperfect information coupled with the desire to set efficiency wages or to avoid the unnecessary costs that would be incurred if information is gained only by observing increased quits.

15. The BLS conducted several special studies of turnover beginning in 1916. Interest in this topic arose from the movement "to promote the more intelligent treatment of laborers" (see Goldberg and Moyes 1985, 98). Various efforts to collect job vacancy statistics were carried out in the 1950s, 1960s, and 1970s. A pilot study for improvement in the late 1970s showed that such an effort would be extremely costly. The then existing labor turnover program was terminated as part of the 1982 budget cut. The Levitan Commission did not recommend the collection of job vacancy statistics. The commission argued that "despite the conceptual appeal of data on job openings, the commission has found job vacancies to be subject to numerous problems of measurement and interpretation, and it doubts that useful data could be obtained at reasonable cost. Aggregate national data would afford little insight beyond what is already provided by other indicators. Vacancy data with occupational, industrial, and area detail would be extremely valuable if combined with similarly detailed data on unemployment and turnover, but the expense of obtaining all these data would be exorbitant" (National Commission on Employment and Unemployment Statistics 1979, 8).

veys of Employer-Provided Training have been conducted. The second survey included interviews with two representative employees from each establishment.

1.1.2 Other Labor Market Information

Labor is a key input into production, and labor income is the major source of personal income. Consequently, it is to be expected that many surveys whose primary focus is not labor markets nonetheless obtain considerable information relevant to them. Major sources of information relating to labor markets collected by other agencies are briefly described in this section.

Household Surveys

Panel Study of Income Dynamics (PSID). The PSID, presently funded by the National Science Foundation and conducted by the University of Michigan Survey Research Center, began in 1968. The initial sample included two subsamples: a probability sample of 2,930 U.S. households and a supplemental sample of 1,872 low-income households taken from the Survey of Economic Opportunity. Individuals from these households have been followed over time with annual interviews. Children born to the sample members become part of the sample. In addition, data have been collected for a supplemental national sample of Latino households beginning in 1990. The PSID has focused primarily on economic well-being and provides a rich body of information including considerable employment-related information. Since the mid-1980s, it has obtained some event history information on employment, income transfers, and demographic states. The PSID also provides detailed information about the neighborhoods (census tracts) in which sample members live.

Survey of Income and Program Participation (SIPP). The SIPP and its predecessor, the Income Survey Development Program data, were developed by the Department of Health and Human Services and the Census Bureau primarily to provide improved and expanded short-term information on participation and turnover in government programs and on sources of income. In the early 1980s, the Census Bureau was given responsibility for the program. SIPP data, collected every four months, provide considerable information on employment and related variables and also on various special topics. There are important differences in concept between the CPS and SIPP unemployment questions, as detailed in Ryscavage and Bregger (1985). A major difference from the CPS is that the SIPP provides the opportunity to examine labor force status *throughout* the month rather than for just the week of the 12th. SIPP data are much less timely than CPS data. SIPP labor force concepts differ from those in the NLS as well. The SIPP has recently undergone a redesign. It will continue with interviews every four months, but now they will continue for four years. In addition, plans call for the Census Bureau to collect data on the 1992–93 SIPP panels annually through 2002. This survey will be called the Survey of Program Dynamics.

Retirement surveys. Among surveys that have supported important work in the labor area are those dealing with retirement-aged individuals. Numerous retirement-related studies were carried out in the past using the NLS older men's survey and the Social Security Administration's Longitudinal Retirement History Survey (LRHS). The LRHS began in 1969 with a sample of household heads aged 58–63. Follow-ups were conducted every two years through 1977. Recognition of the need for new longitudinal information on retirement decisions of older Americans and their well-being after retirement has been widespread; to meet this need, the National Institute on Aging recently began a new Health and Retirement Survey, which is being conducted by the University of Michigan Survey Research Center.

Department of Education longitudinal surveys. Longitudinal studies of young persons sponsored by the Department of Education have been intended to provide information about what occurs at the various levels of education and the major transition phases for students.¹⁶ They have been used for some labor market studies, particularly when detailed information is needed about secondary school(s) attended or about postsecondary schooling of respondents. In particular, the National Longitudinal Study of the Class of 1972 began with interviews of twelfth grade students during the 1971–72 academic year. Follow-ups were conducted in 1973, 1974, 1976, 1979, and 1986. High School and Beyond began in 1980 with high school sophomores as well as seniors. To date, follow-ups have been conducted in 1982, 1984, 1986, and 1992. The National Educational Longitudinal Study of 1988 surveyed eighth grade students; follow-ups were conducted in 1990 and 1992.

Decennial censuses. The decennial Census of Population obtains basic information for the entire U.S. population. Smaller samples of the population have received long-form questionnaires that provide more detailed demographic, socioeconomic, and housing data. The size of the samples has varied over time. Microdata from the decennial censuses are available on various public-use tapes for 1940 and later years. These data have been particularly useful for research studies of immigrants and for studies where a very long time period is of interest, as well as for other purposes where an extremely large sample is needed, such as constructing data for small areas or small demographic groups or for analyses where only weak instruments are available. Planning is under way at the Census Bureau for the 2000 census.

Establishment Surveys

Censuses, annual surveys, and the Longitudinal Research Database (LRD). The Census Bureau conducts a quinquennial Census of Manufactures. The sample for a parallel survey, the Annual Survey of Manufactures (ASM), is

16. The samples are two phased: a sample of schools is selected, then a sample of students within each participating school is interviewed.

drawn from the preceding census. The ASM follows a panel of establishments for five years, beginning in years ending in 4 and 9. Additional establishments are added over the life of the panel to capture births and maintain representativeness. Generally speaking, both the census and ASM contain data on firms with five or more employees. Information is collected on revenues, number of products sold, and costs of about 10 broad classes of inputs, including labor. Two types of employment data are available. Data on total employment and on production worker employment are available for the payroll period including 12 March. Data on production worker employment only are available for payroll periods including 12 February, 12 May, 12 August, and 12 November. Data on the average establishment production worker wage are also available. The Census Bureau, since the 1970s, has developed and maintained a longitudinal file, now called the LRD. The LRD file contains census data beginning in 1963 and ASM data for all manufacturing establishments sampled in the ASMs beginning in 1972. The LRD file has been used for many research studies that could not have been done with other types of data. But in addition to the obvious limitation to only the manufacturing sector, an inherent limitation for some long-term analyses is that the establishments that remain in the sample for more than five years greatly overrepresent large establishments.

Data for service industries are more limited. In 1992, all previously uncovered service industries were added to the economic censuses. Not all industries are presently covered in annual surveys, and content varies by industry. Only two of the annual surveys for services (for SICs 42 and 48) collect information on payrolls; none collect information on employment.

County business patterns. This program provides annual information from Internal Revenue Service administrative records. Outside the Census Bureau, information is available on the number of establishments, employment the week including 12 March, and first quarter and annual payroll for states and counties by two-digit SIC and by more detailed SIC where publication or confidentiality rules are met.

1.1.3 Uses in Labor Economics Research

Labor market information has a variety of important purposes and uses, including academic research studies. Stafford (1986) analyzed the character of labor economics research, the usage of various data sets, and their interaction, for the period 1965–83. In this section, I update his analysis to the period 1984–93. I include articles in the six major U.S. journals he considered.¹⁷ In addition to these “general purpose” journals, I also consider, separately, three

17. The six journals are the *American Economic Review*, *Econometrica*, *Journal of Political Economy*, *Quarterly Journal of Economics*, *Review of Economics and Statistics*, and *International Economic Review*. Following Stafford, I exclude notes, comments, shorter papers, and all articles in the *American Economic Association Proceedings*. Also, separate special supplements are excluded.

major “specialized” journals that focus on labor economics either exclusively or in large part, namely, *Industrial and Labor Relations Review*, *Journal of Human Resources*, and *Journal of Labor Economics*.

In describing the topics of the published papers, I follow the classification scheme used by Stafford, which was based primarily on the chapter topics in Ashenfelter and Layard (1986). Many articles deal with more than one topic category. In this analysis, as in Stafford, each article is assigned to only one category; the assignment is based on the primary focus of the article. Drawing the line between what to count as a labor economics study and what not to count is far from straightforward. I have followed the rule that articles included should deal with labor issues in large part. Thus, for example, a study of economic growth with some reference to human capital would be excluded, as would a theoretical piece on industrial organization with some reference to labor. To be consistent with Stafford and with Hamermesh’s (1993) discussion, I have included general factor demand studies that include estimates of labor demand elasticities or substitution elasticities between labor and other factors of production.¹⁸

This analysis does not attempt to describe the overall importance of various data sets in research; not only are some economics journals excluded from my survey but also many research uses of some of the data sets described above are excluded.¹⁹ For example, macroeconomic studies not focused on labor markets often use the overall unemployment rate from the CPS or employment figures from the CES, but such studies are not included here. As another type of example, many studies using NLS data are conducted by social scientists other than economists or focus on nonlabor topics. To some extent, the present analysis is backward looking in that it does not consider current usage as would be reflected in, say, a review of working papers. However, the focus on major journals has the advantage of limiting studies considered to those that have particular merit as judged by peer review.

Topics

Table 1.2 presents the percentage distribution of articles by major topic area for the period 1984–93, following Stafford’s categorization. For comparative purposes, information for 1965–83 from Stafford is reproduced here. Additional detail is provided in appendix table 1A.1. I have broken out several additional separate subcategories under all but the labor supply category.

The distribution for the general journals that I find for 1984–88 is only somewhat different from that found by Stafford for the immediately preceding period. It is important to note that some portion of the difference across these periods may be due to coding differences rather than to real changes.

18. Hamermesh (1993, 6–7) presents counts of articles in the six journals considered by Stafford for 1984–90 for the labor demand and labor supply categories.

19. In addition, I have not attempted to determine which empirical papers had “meaningful” theoretical sections or to determine which papers were “significant” as Stafford did.

Table 1.2 **Percentage Distribution of Articles by Topic, 1965–93**

Topic	General Journals						Specialty Journals	
	1965–69	1970–74	1975–79	1980–83	1984–88	1989–93	1984–88	1989–93
Labor supply	25.0	32.0	34.0	26.0	23.1	23.1	20.5	23.2
Labor demand	15.5	9.0	8.0	12.0	11.2	12.0	0.9	2.4
Wage determination and earnings	25.5	37.0	34.0	29.5	24.8	28.9	29.1	26.8
Labor market equilibrium and friction	18.0	17.0	18.0	25.0	27.3	25.3	18.7	20.2
Institutional structures	13.0	5.5	5.5	8.0	13.6	10.7	30.9	27.4

Sources: For 1965–83, Stafford (1986, 392); for 1984–93, author's computations.

Continuing to focus on the general journals, note that labor supply accounted for a slightly lower percentage of articles in 1984–93 than Stafford found for 1980–83, which was notably lower than in the preceding 10 years. The percentage of articles on labor demand, small throughout the 1965–83 period, remained about the same.²⁰ The fraction of articles on wage determination and earnings fell somewhat in 1984–88 compared to earlier periods but rose again in 1989–93. Compared to preceding periods, the proportions of articles dealing with labor market equilibrium and friction and with institutional structures rose in 1984–88 but fell back in 1989–93.

The distribution of topics covered in the specialty labor journals was very different from that in the general journals. Articles on labor demand accounted for an extremely small proportion of articles in these journals. The proportion of articles on institutional structures was well over twice as high in the specialty journals as in the general journals, and the proportion on labor market equilibrium and friction was somewhat smaller.

It is interesting to note that only seven papers published during 1984–93 (three of them in general journals) were primarily concerned with data measurement issues. Stafford found only one such article in his period. All measurement papers are classified in the present study according to the subject addressed.

Data Usage

Table 1.3 presents information on the use of various types of data in labor economics articles during the period 1965–93.²¹ The trend toward an increasing proportion of theoretical articles in the general journals stabilized in the 1984–88 period but declined in 1989–93, the first decline in the entire period considered. The proportion of theoretical articles in the labor journals also declined between 1984–88 and 1989–93 and was much lower than in the general journals in both periods. Another striking result concerns the importance of microdata. The percentage of articles based on microdata in the general journals increased in 1984–88 compared to the two preceding periods and increased even more in 1989–93. Microdata usage was even greater in the specialty journals. Table 1.3 reveals a decline in labor economics articles in general journals based on time-series data between 1984–88 and 1989–93, continuing the long-term decline in the use of this type of data in labor economics research.²² The proportion of such articles in the specialty journals was

20. Taking a narrower definition, including only articles with a major focus on labor and thereby excluding general factor demand studies, results in considerably smaller shares of articles devoted to labor demand: 6.9 percent of articles in general journals in 1984–88 and 7.0 percent in 1989–93.

21. Articles are counted as theoretical if they presented no empirical analysis, although they might cite a few previously published figures or pick estimates from previous studies for use in an illustrative simulation. This table counts data set usage, not articles; there are 63 articles that used multiple data sets (some of which fall in the same data type category, however).

22. The exclusion of general factor demand studies, most of which are based on time-series data, results in even lower proportions of articles based on time-series data: 12 percent in 1984–88 and 8 percent in 1989–93.

Table 1.3 **Percentage Distribution of Articles by Data Source, 1965–93**

Data Used	General Journals						Specialty Journals	
	1965–69	1970–74	1975–79	1980–83	1984–88	1989–93	1984–88	1989–93
Theory only	14	19	23	29	29	23	17	10
Microdata	11	27	45	46	51	64	70	82
Time series	42	27	18	16	16	12	8	6
Local area cross section	3	2	4	3	4	3	4	5
State	7	6	3	3	2	1	1	1
Other aggregate cross section	14	16	8	4	2	2	2	1
Secondary data analysis	14	3	3	4	2	1	3	3
Laboratory experiments	– ^a	– ^a	– ^a	– ^a	0	1	0	1
Number of articles	106	191	257	205	242	225	327	332

Sources: For 1965–83, Stafford (1986, 392); for 1984–93, author's computations.

^aStafford did not have a laboratory experiments category.

much lower. Table 1.3 also shows that the percentages of articles based on secondary data and on local area cross-sectional, state, or other aggregate cross-sectional data continued to be extremely low, as they were for the latter part of Stafford's period.

A breakdown of the percentage distribution of labor articles by more detailed categories of data sources is given in table 1.4 for 1984–93. Household or individual survey microdata accounted for by far the largest share of microdata sources. The percentage of articles based on household microdata increased between 1984–88 and 1989–93 as did the percentage of articles based on establishment microdata. The use of microdata from administrative records (not matched to household survey data) was quite low. Use of microdata on other units, primarily unions or bargaining pairs, was fairly low in the general journals but more substantial in the specialty journals.

The important role of the CPS, NLS, and PSID in empirical research in labor economics found by Stafford continued, with use of the CPS increasing somewhat. The use in the general journals of household data from social experiments, which accounted for only 2 percent in 1975–83, increased to 5.4 percent in 1984–88 and 7.6 percent in 1989–93; usage of these data was fairly similar in the specialized journals.

As pointed out forcefully by Hamermesh (1990), microdata for establishments have been lacking in the United States.²³ However, articles using such data did account for 7.7 percent of articles in the general and specialty journals in 1984–88; this rose to 11.0 percent in 1989–93. The establishment microdata sources used the most frequently in these articles were the Employment Opportunities Pilot Project (EOPP) establishment data, 10 articles; Equal Employment Opportunity data and the BLS's Industry Wage Surveys, 6 articles each; and Compustat data and the LRD, 5 articles each. More articles used firm or establishment microdata obtained by the individual researchers themselves than used any one of the above sources.

Of all these labor articles published in 1984–88, 6.7 percent used data from countries other than the United States. This increased to 8.5 percent of articles in 1989–93.

Discussion

To what extent are empirical studies stimulated by theoretical developments and policy interests, and to what extent are they stimulated by data availability? Stafford argued that “the greater range and quality of our knowledge and theorizing on particular topics does significantly stem from the advent of large-scale micro level datasets. These datasets use individuals and households as the unit of observation and are important in explaining why, during the 1970s,

23. As can be seen from references in Hamermesh (1993) to labor demand studies using establishment data from other countries, the United States has lagged behind a number of other countries in use of establishment microdata.

Table 1.4 Percentage Distribution of Articles by Detailed Data Type, 1984–93

Data Used	General		Specialty		All Journals	
	1984–88	1989–93	1984–88	1989–93	1984–88	1989–93
<i>Theory only</i>	28.9	22.7	16.8	9.9	22.0	15.1
<i>Household survey microdata</i>	40.9	47.6	49.8	59.6	46.0	54.8
PSID	7.9	7.1	6.7	6.0	7.2	6.5
NLS	4.1	7.1	7.0	7.8	5.8	7.5
CPS	8.7	8.4	11.6	15.4	10.4	12.6
SIPP	0.0	2.2	0.0	0.9	0.0	1.4
Department of Education	0.4	0.0	0.9	1.8	0.7	1.1
Decennial census	1.2	1.8	2.1	2.4	1.8	2.2
Other household data	16.9	20.9	19.6	24.1	18.5	22.8
Other international household data	4.5	2.2	0.9	1.5	2.5	1.8
Other social experiment data	5.4	7.6	4.6	9.0	4.9	8.4
Household/administrative matches	1.7	0.0	1.8	1.2	1.8	0.7
<i>Firm or establishment survey microdata</i>	6.2	9.8	8.9	11.4	7.7	10.8
EOPP (establishment only)	0.4	1.3	0.6	1.2	0.5	1.3
LRD	0.0	1.8	0.0	0.3	0.0	0.9
Other establishment data	5.8	6.7	8.3	9.9	7.2	8.6
Other international establishment data	0.4	0.9	0.6	1.8	0.5	1.4
<i>Microdata from administrative records</i>	1.7	2.7	3.1	2.7	2.5	2.7
Individual microdata	1.7	2.7	1.8	1.2	1.8	1.8

Individual international microdata	0.0	0.0	0.6	0.3	0.4	0.2
Establishment microdata	0.0	0.0	0.6	0.9	0.4	0.5
Establishment international microdata	0.0	0.0	0.0	0.3	0.0	0.2
<i>Other microdata</i>	2.5	3.6	8.0	7.5	5.6	5.9
Other microdata	0.0	0.9	0.3	0.3	0.2	0.5
Other international microdata	0.0	0.0	0.3	0.0	0.2	0.0
Labor contract/strike microdata	2.5	2.7	7.3	7.2	5.3	5.4
BLS Compensation and Working Conditions	1.2	0.9	2.4	3.0	1.9	2.2
Other international contract data	0.4	1.8	1.5	1.8	1.1	1.8
<i>Time series</i>	16.1	11.6	8.3	6.0	11.6	8.3
Time-series data	14.5	8.0	7.6	4.2	10.5	5.7
International time-series data	1.7	3.6	0.6	1.8	1.1	2.5
<i>Local area cross section</i>	3.7	3.1	4.3	5.4	4.0	4.5
Census tract	2.1	3.1	1.2	3.0	1.6	3.1
International census	0.8	0.0	0.3	0.3	0.5	0.2
Other area cross section	0.8	0.0	2.8	2.1	1.9	1.3
<i>State</i>	1.7	0.9	0.9	0.9	1.2	0.9
<i>Other aggregate cross section</i>	2.1	1.8	1.8	1.5	1.9	1.6
Other aggregate cross section	1.2	1.3	1.8	1.2	1.6	1.3
International aggregate cross section	0.8	0.4	0.0	0.3	0.4	0.4
<i>Secondary data analysis</i>	2.1	0.9	3.4	3.0	2.8	2.2
<i>Experimental data</i>	0.4	0.9	0.0	1.2	0.2	1.1
<i>Number of observations</i>	242	225	327	332	569	557

about two-thirds of labor articles in major journals were on the broad subjects of labor supply and wage determination" (1986, 388). Hamermesh (1993) argues that the study of labor demand has been conditioned by the availability of data. At the same time, however, data collection itself necessarily is conditioned by demand for it. While the demand for data is primarily affected by the needs of government policymakers and the business community, it is likely to be indirectly responsive to research developments.

The decline in the share of theoretical articles in 1989–93 in labor economics research does not correspond to the appearance of whole new government data sources. There were not many new sources, and such sources were little used (i.e., SIPP, although all of its use came in the most recent period 1989–93). The Quality of Employment Survey, which interviewed workers in 1972–73 and 1977 but has not been continued, was used in ten papers in 1984–88 and only two in 1989–93. Data from the social experiments that were used most frequently during 1984–93 (notably, the Seattle Income Maintenance Experiment/Denver Income Maintenance Experiment and the EOPP) are becoming outdated; recent experiments have been more specialized, and I expect use of this type of data in labor economics to decline in the near future.

In discussing the large volume of research on labor supply and related issues that was produced over the period 1965–83, Stafford (1986) specifically mentioned the rich microdata that became available for the first time early in that period from the CPS, NLS, and PSID. Stafford expected that continued use of these data sources might be primarily a reworking of the same topics, but that has not been the case. Situations facing workers have changed so that analyses using data for recent years provide additional information far beyond reworking.²⁴ New uses of the NLS and PSID have arisen since the early 1980s. In an area where research seems to be expanding, both the NLS and the PSID provide excellent opportunities for studies of intergenerational or within-family linkages of welfare dependency and income and employment. Additional topics have been covered in these surveys in recent years, and this expanding or revised content has been important in the most recent 10 years, as Stafford (1986, 404) suggested was the case for his period. In addition, the development of event history data permits examination of new aspects of the labor market experiences of workers in addition to permitting improved analyses of topics previously addressed. The increased use of the CPS is not primarily due to expanding content. CPS content has been expanded or revised largely through additional special supplements. One new supplement, the Displaced Workers Supplement, accounted for just under 1 percent of the labor articles. Increased accessibility of CPS microdata may have been responsible for its increased use.

24. The view that analyses can be done with very old data and provide still relevant information often seems implicit in microeconomic research. This is in stark contrast with the business and policy communities, which find anything more than a few years old, or in some cases a few months old, obsolete and uninteresting.

Determining the extent to which theoretical studies and policy interests are stimulated by data availability and vice versa is beyond the scope of this paper. Tabulations presented in table 1.5, which include articles in both the general and specialty journals, are useful in assessing the much more limited question of the extent to which the focus on broad topic areas differed between theoretical and empirical articles during 1984–93. It is interesting to note first that the percentage of articles that are theoretical varies considerably across major topic areas, from a high of 34 percent in the labor market equilibrium and friction category to a low of 10 percent in the wage determination and earnings category. Within some of these topic categories there is also considerable variation in the percentage of articles that are theoretical, most notably in the labor market equilibrium and friction category. Sixty-five percent of the articles on screening, signaling, contracting, and matching are theoretical. In contrast, only 6 percent of articles on career mobility, gross job flows, and other turnover are theoretical.

Among empirical articles, the various types of data are used to very different extents in studying different topics. Household microdata are used in the majority of studies of labor supply and of wage determination and earnings and in 38 percent of the studies of labor market equilibrium and friction (which account for well over half of the empirical articles in this latter topic area). As already noted, time-series data are used far more extensively in analyses of labor demand than in the other areas. Only a small fraction (11 percent) of the use of establishment microdata is in studies of labor demand, a very small category; these data are used in about 17 percent of labor demand studies. In fact, the majority of articles on all the major topics except labor demand use some type of microdata. Use of local area cross-sectional data, state data, other aggregate cross-sectional data, and secondary data is not quantitatively important for any of the topics.

This analysis has shown that empirical economics was alive and well in labor economics research over the 1984–93 period. Unlike the 1969–83 period analyzed by Stafford, this was not primarily due to development of new government survey efforts. Although the improvement of econometric methods played a role in this later period as in the earlier years, and the expanded content of ongoing surveys was also important, the dramatically reduced cost of computing may have played an important role. The reduction in computational costs has not only made the use of complex econometric methods feasible but has also permitted many studies that use great amounts of data (such as many years of CPS microdata) while not exploiting new methods.

The ultimate benefit from empirical research is not the number of articles published but the extent to which understanding of the economy is improved, policy discussions are informed, and programs can be evaluated. Topel's paper (chap. 2) in this volume assesses data needs and success of labor economics studies for two purposes: description and estimation of the parameters of economic models. Section 1.2 of this paper addresses needs for data, including uses other than in research studies.

Table 1.5 Cross-Tabulation of Major Topic by Data Type, 1984–93

Data Used	Labor Supply	Labor Demand	Wage Determination and Earnings	Labor Market Equilibrium and Friction	Institutional Structures	Total
Theory only	36.0	11.0	30.0	85.0	47.0	209
Household microdata	183.5	8.0	185.3	95.5	54.0	526
Establishment microdata	2.0	11.0	28.7	27.5	27.5	97
Administrative microdata	6.0	0.0	8.5	9.0	2.5	26
Other microdata	1.0	0.0	3.5	3.0	52.8	60
Time series	10.5	28.0	19.0	16.3	32.8	107
Local area cross section	6.0	3.0	15.0	2.0	16.0	42
State	3.0	1.0	2.0	2.8	1.5	10
Other aggregate cross section	1.0	3.0	4.0	5.5	3.3	17
Secondary data	3.0	0.0	11.0	4.3	7.0	25
Experimental data	0.0	0.0	1.0	0.0	5.5	7
Total	252	65	308	251	250	1,126
	<i>Column Percentages</i>					
Theory only	14.3	16.9	9.7	33.9	18.8	
Household microdata	72.8	12.3	60.2	38.1	21.6	
Establishment microdata	0.8	16.9	9.3	11.0	11.0	
Administrative microdata	2.4	0.0	2.8	3.6	1.0	

Other microdata	0.4	0.0	1.1	1.2	21.1	
Time series	4.2	43.1	6.2	6.5	13.1	
Local area cross section	2.4	4.6	4.9	0.8	6.4	
State	1.2	1.5	0.6	1.1	0.6	
Other aggregate cross section	0.4	4.6	1.3	2.2	1.3	
Secondary data	1.2	0.0	3.6	1.7	2.8	
Experimental data	0.0	0.0	0.3	0.0	2.2	
Total	100	100	100	100	100	
		<i>Row Percentages</i>				
Theory only	17.2	5.3	14.4	40.7	22.5	100
Household microdata	34.9	1.5	35.2	18.1	10.3	100
Establishment microdata	2.1	11.4	29.7	28.4	28.4	100
Administrative microdata	23.1	0.0	32.7	34.6	9.6	100
Other microdata	1.7	0.0	5.8	5.0	87.6	100
Time series	9.8	26.3	17.8	15.3	30.8	100
Local area cross section	14.3	7.1	35.7	4.8	38.1	100
State	29.1	9.7	19.4	27.2	14.6	100
Other aggregate cross section	5.9	17.8	23.8	32.7	19.8	100
Secondary data	11.9	0.0	43.5	17.0	27.7	100
Experimental data	0.0	0.0	15.4	0.0	84.6	100

Note: If an article uses n data sets, each use is counted as $(1/n)$ th of a data set.

1.2 Other Uses, Needs, and Possibilities

In their examination of the need to improve federal statistical programs, Duncan and Gross suggest that “as the 21st century approaches, it is increasingly clear that our conceptual net, designed for earlier realities, fits our current situation only very imperfectly” (1993, 11). Is this true of labor market data? What seems clear on the surface is that the system is fragmented and certainly is not what would be designed for research or policy purposes if one were starting from scratch in an ideal world. But in a world of limited resources, this is less clear: each survey has users who consider it important. Timeliness, periodicity, and detail differ among surveys, and needs for them differ by purpose.

Of course, a particular measure cannot serve all uses and users. Sometimes a finding that a particular measure is inappropriate for a certain purpose leads to development of a new survey. For instance, the ECI was developed in response to a call for a suitable measure of labor costs. But some of the information collected separately in the past might have been developed instead through a consolidation of programs.

What are the needs and how do we determine them? Ultimately, the demands of economic policymakers and private sector users will influence this, as was certainly the case in the early years of the BLS. But it has been argued by Triplett that “today’s research need becomes tomorrow’s policy-analytic need” (1990, 343). This suggests a key role for researchers in helping to inform statistical agency concerns. As Triplett points out, historically researchers have had little ability to provide input into agency decisions. In fact, some efforts have been driven by research and policy analytic needs outside of statistical agencies—as the major examples in labor economics, household longitudinal surveys were not begun by statistical agencies, although the NLS and SIPP have been transferred to them. At the beginning of her tenure at the BLS in 1993, Commissioner Abraham announced two major goals: improving access to data, particularly microdata, and expanding survey content. This provides perhaps an unparalleled opportunity for research and researchers to help shape statistical programs.

This part of the paper begins by addressing recognized needs for additional information. Resources for maintaining the statistical system, let alone expanding the data provided, have been restrained over the past decade. In view of this, the paper concludes with a discussion of how to make the system more useful in a situation where resources are very tight.

1.2.1 What Is Missing? Recognized Needs

Content

The discussion of existing programs suggests that new or changed data programs frequently arose from the interests of the day. It does not appear that

there have been many concerted calls for additional information that have gone unrecognized over the past 10 or 15 years, at least not for things that would call for whole new programs.

Some emerging issues can be addressed with existing data, although perhaps imperfectly. The major research topics of declining real wages and increasing wage dispersion have perhaps been driven by the recognition of trends in existing data and so far do not appear to call for more data. But there are related concerns as well as other issues for which data do not exist or on which existing data are extremely limited.

Labor supply behavior of youth and retirement decisions. During the mid- to late 1980s it became clear that in spite of the existence of major household surveys there were important gaps in data for analyzing key labor supply phenomena. Because the experience of existing NLS sample members during their youth is not expected to be representative of the experience of today's youth who face very different labor market and family situations, a primary need was for development of a new survey of youth. This need will be satisfied by beginning a new NLS youth cohort; this is probably best viewed as a sample replenishment effort rather than a "new" effort. The other major area of need was for development of data suitable for analyzing the current retirement behavior of Americans. This need has been met by the National Institute on Aging's new Health and Retirement Survey.

Contracting out and contingent work. Since the mid- to late 1980s, there has been recognition of changing types of employment, in terms of the growth of contracting out and of "contingent work." Some information on the former was provided through special surveys. Information on contingent workers is being collected in a special supplement to the CPS and also in the NLS.

Nonwage compensation. Data on compensation as opposed to money wages have been very limited. This may have particularly important implications for understanding labor markets because compensation practices may be continually changing. Even the definition of cash wages is problematic and differs across surveys. In the mid-1980s, there was attention at the BLS to the increasing use of lump-sum payments in lieu of wages and how they should be treated in series that include them. While it had appeared that this trend was stabilizing or declining, conversation with industry specialists and some recent union agreements (e.g., the recent agreement at the Xerox Corporation facility in Webster, New York; *Washington Post*, 9 June 1994) suggest that use of some type of bonuses or payments in lieu of wage increases may be reemerging as a trend, implying a need for broader measures of cash compensation.

Labor demand. Hamermesh (1990, 1993) and others have called for improved data for studying labor demand issues. Hamermesh specifically recommended

a relatively small but representative quarterly or monthly survey of establishments with extensive content: employment by major skill category, hours worked by each group of workers, payroll for each type of worker, other labor costs, and total sales and production. This he suggested would be an extension or rationalization of existing information. Hamermesh would ideally also include some key information on worker characteristics through a linked survey. Like Hamermesh, Juster (1983) called for more information relating to the demand for labor but recommended information on hiring practices, on-the-job training, and how firms evaluate worker productivity and decide on promotions, raises, firings, and so forth.

A related point is the recognition by a variety of data users that there has been no detailed establishment information on wages by occupation for all industries. The lack of such information has been noted frequently by analysts, and sometimes by users who would want the microdata together with information on other costs faced by establishments. Thus analyses of wages by occupation and industry typically have been carried out with the CPS, for which there is concern about coding of industry and occupation and which does not provide other establishment information. The planned OES expansion will address this need.

Training and workplace practices. Training and workplace practices and their relation to wages, employment growth, and productivity are issues of considerable current policy concern. Detail on the structure of compensation is needed to test alternative theories about work organization and incentives as well as for completeness to avoid erroneous conclusions about what is happening to employee well-being. Additional information needed includes measures of amount of supervision and worker control over the job, hiring and promotion practices, and various characteristics of the workplace, including use of teams or quality circles. As already noted, a small establishment Survey of Employer-Provided Training with worker information was conducted by the BLS in 1995.

Target Population

It is interesting that in the United Kingdom as well as in the United States early collection of data focused on certain types of workers—"manual" workers in the United Kingdom and production workers in the United States—and on manufacturing industries. Perhaps this was true in most countries. In important instances, U.S. establishment surveys still do not provide complete coverage of workers. In addition to sometimes focusing on only production and nonsupervisory workers, U.S. establishment surveys occasionally exclude workers in small firms and typically exclude the self-employed. This latter omission, as noted by Orchard and Stibbard (1993), affects different sectors of the economy differently; notably, certain service industries will be greatly

underrepresented. For understanding the labor market situation, it is critical to have at least some data that cover all of these workers. For instance, major policy issues concern things like whether real earnings are declining and what is happening to low-paid workers relative to the more highly paid. It is important to have complete coverage to satisfactorily address these questions, to make comparisons of wages across industries, and so forth. One advantage of household data is broader coverage.

Household Longitudinal Data

The Levitan Commission pointed to a lack of information on the dynamics of labor market experience. The commission members argued that changes during the 1960s and 1970s gave workers wider choices and that “the data on current labor force status must be supplemented with information on how that status came about and the conditions under which it might change” (National Commission on Employment and Unemployment Statistics 1979, 1). They explicitly recommended the development of publishable gross flow data using the monthly CPS, which would show the numbers of unemployed who remain unemployed, find employment, or leave the labor force in the next month, and similar numbers for those employed or out of the labor force in the first month. Such information has not been published by the BLS because of problems with quality due especially to failure to follow movers and classification error. No adjustment method has been widely agreed to, and there is need for further research. But some argue that such information can help in the interpretation of the monthly labor force situation, despite the measurement problems with existing tabulations (see Barkume and Horvath 1995). If gross flow data could be published, policy interest would likely be high. Currently, the BLS is planning to follow movers in the CPS for one month, which would permit construction of improved gross flow estimates and improved analyses using CPS microdata.

The Levitan Commission also made recommendations for more complete identification of the labor market problems experienced by women and supported special longitudinal studies of women’s lifetime work experience. This need is being satisfied by the continuation and aging of the NLS women’s data together with the initiation of a new youth cohort. Changes in the labor market such as a decline in labor force participation of middle-aged and older men and worsening job prospects for many young men suggest that long-term studies of labor market experience are now essential for men also.

Establishment Surveys with Demographic Data and Linked Household-Establishment Surveys

Linked household-establishment surveys, which have rarely been done in the United States, and to my knowledge are not done in the United States on an ongoing basis, could provide extremely rich information for labor market

analysis.²⁵ Such data were recommended by Hamermesh (1990), and Rosen (1990) argues that they are essential for understanding the matching aspects of labor market exchange.²⁶

Merged establishment-household data are desirable for reasons other than microdata research needs. As is widely recognized, some information can be collected more accurately from establishment surveys—for instance, the industry and occupation (type of work) of the job. But in the United States, if we want to know anything about the human capital or demographic characteristics of workers in an industry or occupation, we almost always have to turn to household surveys. If some information is available only on household surveys and other information is available on establishment surveys, one can bring them together based on tabulations of data from the two sources—that is, one can use data on earnings in an industry from an establishment survey and data on the educational attainment of workers in that industry from a household survey—but since the industry variables are measured differently such unions are less than ideal. It is possible to collect information from establishments on a limited set of demographic variables, as was done, for instance, in the special demographic supplement to the BLS's White Collar Pay Survey used by Bronars and Famulari (chap. 13 in this volume).²⁷ But to collect detailed information on human capital, family characteristics, and nonearnings income, it would be necessary to interview workers. It should be noted however that it would be difficult to develop useful longitudinal worker-employer data because of frequent job changes.

There are other ways of obtaining establishment information linked to some information on workers. Both Orchard and Stibbard (1993) and Hostrup-Pedersen (1993) describe establishment data sets that provide some information on persons, for the United Kingdom and Denmark, respectively. Hostrup-Pedersen describes the use by Danmarks Statistics of register-based data to produce employment data. Detailed tabulations by area are possible, and the distribution of employees within an industry by occupation, education, sex, and age is produced, something that in the United States is regularly available only from household interviews. Denmark's use of a register provides a

25. There have been special linked surveys. Troske uses matched LRD-Census of Population data (chap. 11 in this volume). The EOPP, a test of a proposed welfare reform option, included collection of data for a sample of households and employers in test and comparison sites in 1980. A second wave of the EOPP employer survey, sponsored by the National Institute of Education and the National Center for Research in Vocational Education, provides data for 1981. The BLS is planning a small test to obtain information from a sample of workers within the approximately 1,200 establishments to be surveyed in a second Survey of Employer-Provided Training.

26. In contrast, Rosen was not enthusiastic about Hamermesh's recommendation for additional high-frequency establishment data for the particular purpose of studying labor demand.

27. That effort showed that some kinds of worker information are much easier to collect in establishment surveys than other kinds. Most notably, reporting of information on the starting wage was very low. However, obtaining retrospective information of any type in establishment or household surveys is problematic, which is why longitudinal information surveys are needed for many purposes.

straightforward method for obtaining linked employer-employee information with a frame that should be of extremely high quality. The United States does not have such a register of persons that is used in this way. But as the Orchard-Stibbard paper shows, it is not necessary to have such an individual register to get this information—it can be obtained by beginning with any administrative data set or sample frame providing information on persons. For example, the U.K. New Earnings Survey contains information from employers' payroll records on earnings, sex, age, occupation, industry, and collective bargaining coverage; occasionally, special questions have been included. An important limitation is that this U.K. source does not cover persons below a certain income tax threshold. It is important to note that microdata from these surveys for Denmark and the United Kingdom are not provided to outsiders for research purposes, and it appears they are not used internally for such purposes either.

Some kinds of information we may want for household-based analyses cannot be obtained directly from households. For instance, information on the amount of fringe benefits paid to particular individuals together with information on household demographics and income sources is not available in the United States, and high-quality information of this type would only be possible with a linked household-establishment survey. Thus, whether beginning from the perspective of looking at establishments or households, there are many reasons why linked establishment-worker data are desirable.

1.2.2 How to Make the Existing System More Useful

Microdata Usage and Linkages

An important need recognized by many researchers is increased microdata availability from both household and establishment surveys. In the case of household surveys, there is considerable demand for small-area data and for matched administrative records data. Both are problematic for statistical agencies because of concerns about privacy (see also the discussion below of confidentiality in the context of establishment surveys). Releasing data with information for very small geographic areas greatly increases the probability that a respondent could be identified.²⁸ Matching to administrative records can be particularly problematic. For instance, for the 1990 interview of the NLS older men, waiver forms were developed by the Census Bureau to obtain consent from respondents for matching to social security records but they were not approved in the Office of Management and Budget clearance process.

In contrast to the large volume of research using household microdata, there has been relatively little research using establishment microdata. Use of estab-

28. The PSID provides census tract information, but the NLS has not done so under its geocode agreement. However, recently the NLS has decided to make available local area information from the NLS youth cohort to researchers working at NORC in Chicago; presumably this could be done at the BLS in Washington, D.C., as well.

lishment microdata, both cross sectional and longitudinal, is essential in order to expand the issues that can be addressed. For instance, using LRD data Davis and Haltiwanger (1990) found important results about expansion and contraction of establishments over the business cycle, results that could never have been obtained with aggregate U.S. statistics as previously conceived. Analyses using establishment microdata are needed also to address topics previously studied using aggregated data. As Caballero (1992) has pointed out, fallacies can arise from application of the representative agent framework to model macroeconomic data; to avoid such fallacies of composition, analyses using firm or establishment microdata are required.

At the BLS, establishment microdata have not been kept in the appropriate form for ready use within the agency. Microdata from some surveys have not been saved at all in some instances. In other instances, some data elements collected have been discarded when they are no longer needed for program purposes. The BLS is currently developing longitudinal microdata from the ES-202 program and is forming a CES-linked longitudinal microdata file of respondents from 1972 forward.

While the Census Bureau has for a number of years supported development of the LRD file for research use within the agency, it has been argued that even Census Bureau support for microdata has been insufficient (see McGuckin 1991). But it must be noted that uses are also very limited as long as no way is established to make these data widely accessible to outside researchers. In fact, my initial view of Hamermesh's (1990) call for a whole new effort on labor demand was that it was hard to see how it could be justified unless some way could be found to improve access to the data.

For the past 10 years or so, outside researchers have been able to access confidential microdata while serving as ASA/NSF fellows at either the Census Bureau or the BLS. Some other researchers have also been able to access the LRD data at the Census Bureau, and a few outside researchers have used confidential microdata at the BLS. The Census Bureau has recently established a regional data center. Considerable thought is now being given at the BLS to making data more accessible while preserving confidentiality. In addition to expanding possibilities for use of these data within the agency by both BLS and outside researchers, another approach might be through use of special agreements, as is done by the BLS for the NLS youth cohort geocode data. An alternative approach, masking data, has been discussed in various contexts for a long time; this is not a promising approach because it seems unlikely that masked data could meet very many needs, given the wide range of interests of researchers and the different econometric issues that are pertinent for each study.

Other countries too have been slow to make available establishment microdata for research purposes. Among microdata that have been used for research are French data used to study investment and labor demand (Mairesse and Dormont 1985) and compensation (Abowd and Kramarz, chap. 10 in this volume).

In Great Britain, there has been one major establishment survey for which microdata are publically available for research, the Workplace Industrial Relations Survey (WIRS), begun in 1980 (see, e.g., Millward 1993 for a discussion). According to Hamermesh, the WIRS, which focuses on the practices of management-employee relations at the workplace, provides “most of the variables one might want for studying labor demand” (1993, 398). It has been collected and processed by a private research organization rather than a government agency, although three government agencies, including the Department of Employment, have been sponsors. Research topics studied using this survey have included the impact of unions of various types on wages and the effects of profit-related pay. Similar surveys have recently begun in Australia and France.

The Census Bureau has increased the number of issues that can be addressed using the LRD file by linking it with other surveys—for instance, the wage data described in the paper by Troske (chap. 11 in this volume). There are gains possible from matching BLS files as well. In the past, limited efforts have been made by researchers within the BLS: Pergamit (1987) matched CES data with the union status variable from the Area Wage Surveys for two states. Ruser (1993) matched Occupational Safety and Health Statistics annual data on injury counts, hours, and employment to CES data on the weekly wage, percentage of production workers, percentage female, and weekly overtime hours. But such efforts have been difficult for a single researcher, and many possibilities have been prevented or made extremely costly by data’s not being stored and maintained in a suitable way. In theory, matching at least some historical data is possible fairly readily across all BLS establishment surveys. Unfortunately, such matches (except those with the ES-202) will only yield coverage of large firms since there will be very few matches of smaller firms that are sampled with low probability. Matching “special surveys” to ongoing surveys may also be beneficial if enough matches could be obtained. For instance, matching a special survey of workplace practices, if one were to be conducted, to ECI and EBS data would greatly enhance research possibilities because of the importance of information on compensation structure. It is important to note the problem of sample representativeness when samples are linked over time and when differing surveys are matched, but regression studies can still be performed and statistical work to figure out how to create improved estimates from such matches might be possible.

In our decentralized statistical system, matching confidential data collected by different agencies is not possible at present. Important improvements in knowledge about the U.S. economy would come about through matching the Census Bureau’s LRD file with BLS data, for instance, producer price index data or occupational safety and health data. This would necessitate data-sharing legislation or some other mechanism. However, as with matching existing BLS surveys, there would be a problem of few matches except for large establishments.

Developing and maintaining readily accessible longitudinal microdata files

seems a minimal part of any effort to improve establishment data for research purposes. Development of matched files across surveys may also be valuable. An advantage is that, at extremely low cost, knowledge can be expanded. It might also be possible to consider changes in the surveys themselves that would improve research uses while continuing to permit historic program uses; such changes may not always be particularly expensive and may even save money. An example is that, starting in 1989, the same establishments have been included in the ECI and EBS. While this was done for cost reasons and the two surveys have not been processed together, the common sample will provide interesting research opportunities. A conference of researchers was held in September 1994 to address possibilities for a microdata file containing both sets of information—among the difficulties in this are differing units of observation, that is, occupation in the ECI and benefit plan in the EBS.

Improving Content

One of Commissioner Abraham's major goals is to expand the content of existing surveys. In the recent past, relatively little new information has been added to or proposed for existing surveys.

The American Statistical Association Panel report (1994, 10, 30) recommended investigating the feasibility of collecting additional data through the CES survey.²⁹ If a probability sample is implemented, the addition of information to the CES would be an obvious place to begin developing information for analyzing labor demand. For such a huge sample, adding information to only a small subsample of the monthly survey or fielding a special supplement quarterly or annually would provide a basis for numerous analyses. Presumably, it would not be feasible to collect nearly as extensive a set of information as would be ideal for some uses—see Hamermesh (1990, 1993)—because it is not a personal interview survey. In fact, respondent burden would probably be seen as limiting how much desirable information would be collected even in a personal visit survey.

New Uses

In the current budget environment, it appears that not very many major new efforts will be feasible. Therefore, improvements in use of existing data that can be achieved at modest cost are an important focus. Development of gross change measures from both household and establishment data is an important need. As discussed above, Davis and Haltiwanger's work on gross employment change provides major new measures for the United States but refers only to the manufacturing sector. Work at the BLS is just beginning to construct simi-

29. The American Statistical Association Panel report also recommends research on "the development of CES concepts that would more nearly meet the needs of the user community," including "where needed, to investigate the feasibility of changing the set of variables collected . . . to meet users' needs" (1994, 6). The panel recommends consideration of obtaining data on broad occupational breakdowns and on lump-sum payments and employee contributions to thrift savings plans.

lar measures using ES-202 data that will include all industries and be representative of small firms. Also, measures of gross change in payrolls and hours might be constructed using the CES data. Adding content to existing surveys, discussed above, is another way of obtaining new information at relatively low cost.

Current research by Anderson and Meyer (1994) using individual microdata from UI administrative wage records for eight states for 1978–84 combines information on individual workers and establishments. The view taken is an establishment perspective. They divide turnover into temporary separations and accessions, permanent job position creation and destruction as firms grow or decline, and job match creation and destruction as individuals move across continuing positions at different firms. Since little demographic information is available, and hours data are lacking, these data cannot be used to address many of the issues that researchers would want to explore using an ideal matched firm-worker data set. But clearly Anderson and Meyer's analysis extends our knowledge in important ways and suggests additional statistics that could be produced with existing information. A possible new effort is being discussed at the BLS to create a wage record microdatabase. If these data could be linked to CPS microdata, the research payoff could be enormous, both in terms of assessing the quality of earnings data obtained from both sources and for providing demographic information for a variety of analytic purposes, including expanding on the type of analyses done by Anderson and Meyer. But data-sharing legislation would clearly be needed to enable this. NLS youth cohort data might also be linked to UI data.

Trade-Offs

The focus in this paper has been on research needs, but of course there are many other major uses for labor statistics. Research needs cannot be addressed apart from these other needs in a tight budgetary period. There is considerable interest on the part of a variety of users, including policymakers, the Federal Reserve Board, macroeconomic forecasters, and financial sector analysts, in very timely monthly aggregate statistical series.

Another major use of BLS data is in automatic formulas for indexing and other purposes. In the case of labor measures, ECI data are well suited to use in escalation where a measure of labor costs is needed. Among such uses, the Health Care Financing Administration uses the ECI as part of an input price index for prospective reimbursement of hospital charges under Medicare. Average hourly earnings is also used for escalation purposes in private sector contracts because of its extensive industry detail. In addition, local-area unemployment statistics (constructed using the CPS and statistical methods) have been used for allocating funds under the Comprehensive Employment and Training Act and successor programs.

Another need is for large sample sizes to provide state and area data and more statistical precision. For instance, there is demand by states for data to

use in forecasting revenues. But there is a trade-off between this need and the need for richer information.³⁰

Trade-offs clearly exist between features of programs that serve needs for timely macrodata, indexing or allocating funds, and state and area detail and features of programs that serve research and other policy needs. Trade-offs exist between accuracy, frequency, and detail of information provided. Consideration of these sorts of trade-offs is beyond the scope of this paper.

30. For instance, in discussing Antos's (1983) paper, which described BLS data on compensation, Juster (1983) recognized the desires of others such as states for detailed data but called for a different trade-off. He stated: "I would make the judgment that the BLS program would be better if it collected data from fewer units of observation generally, but measured more variables for those same units. The basic reason is that I would be prepared to live with somewhat more sampling variance, if the benefits were a substantially enhanced analytic potential" (180).

Appendix

Table 1A.1 **Percentage Distribution of Articles by Detailed Topic, 1984–93**

Topic	General Journals		Specialty Journals		All Journals	
	1984–88	1989–93	1984–88	1989–93	1984–88	1989–93
<i>Labor supply</i>	23.1	23.1	20.5	23.2	21.6	23.2
Population size and structure	2.5	1.3	0.6	0.9	1.4	1.1
Household production	1.2	3.6	0.9	2.4	1.1	2.9
Labor supply of men	0.0	0.9	2.1	0.9	1.2	0.9
Labor supply of women	2.1	2.7	2.1	3.6	2.1	3.2
Other labor supply	9.9	8.4	6.1	9.3	7.7	9.0
Retirement	2.1	1.3	4.9	2.7	3.7	2.2
Education demand	1.7	2.2	1.5	1.5	1.6	1.8
Migration	3.7	2.7	2.1	1.8	2.8	2.2
<i>Labor demand</i>	11.2	12.0	0.9	2.4	5.3	6.3
Basic labor demand	4.5	4.0	0.6	0.9	2.3	2.2
Adjustment and dynamic demand	0.8	2.2	0.3	0.0	0.5	0.9
Minimum wage	1.2	0.4	0.0	1.5	0.5	1.1
Static factor demand	4.1	3.1	0.0	0.0	1.8	1.3
Dynamic factor demand	0.4	2.2	0.0	0.0	0.2	0.9
<i>Wage determination and earnings</i>	24.8	28.9	29.1	26.8	27.2	27.6
Earnings functions	12.0	12.0	8.9	13.3	10.2	12.7
Theoretical lifetime earnings	0.0	0.9	0.3	0.0	0.2	0.4
Compensating wage differentials	2.1	1.3	0.9	0.9	1.4	1.1

(continued)

References

- American Statistical Association Panel. 1994. A research agenda to guide and improve the Current Employment Statistics survey. Alexandria, Va.: American Statistical Association, January.
- Anderson, Patricia M., and Bruce D. Meyer. 1994. The extent and consequences of job turnover. *Brookings Papers on Economic Activity: Microeconomics*, 177–237.
- Antos, Joseph R. 1983. Analysis of labor cost: Data concepts and sources. In *The measurement of labor cost*, ed. J. E. Triplett. Chicago: University of Chicago Press.
- Ashenfelter, Orley C., and Richard Layard, eds. 1986. *Handbook of labor economics*. New York: Elsevier.
- Barkume, Anthony J., and Francis W. Horvath. 1995. Using gross flows to explore movements in the labor force. *Monthly Labor Review* 118 (April): 28–35.
- Caballero, Ricardo J. 1992. A fallacy of composition. *American Economic Review* 82 (December): 1279–92.
- Davis, Steven, and John Haltiwanger. 1990. Gross job creation and destruction: Microeconomic evidence and macroeconomic implications. In *NBER macroeconomics annual 1990*, ed. O. J. Blanchard and S. Fischer, 123–86. Cambridge, Mass.: MIT Press.
- Duncan, Joseph W., and Andrew C. Gross. 1993. *Statistics for the 21st century*. New York: Dun and Bradstreet.
- Goldberg, Joseph P., and William T. Moye. 1985. *The first hundred years of the Bureau of Labor Statistics*. BLS Bulletin no. 2235. Washington, D.C.: U.S. Department of Labor, Bureau of Labor Statistics, September.
- Hamermesh, Daniel S. 1990. Data difficulties in labor economics. In ed. E. R. Berndt and J. E. Triplett, 273–95. *Fifty years of economic measurement*, Chicago: University of Chicago Press.
- . 1993. *Labor demand*. Princeton, N.J.: Princeton University Press.
- Hostrup-Pedersen, Soren. 1993. Statistics of employment in businesses—Denmark. In *Proceedings of the International Conference on Establishment Surveys*, 309–15. Alexandria, Va.: American Statistical Association.
- Juster, F. Thomas. 1983. Comment. In *The measurement of labor cost*, ed. J. E. Triplett. Chicago: University of Chicago Press.
- Mairesse, Jacques, and Brigitte Dormont. 1985. Labor and investment demand at the firm level: A comparison of French, German and U.S. manufacturing, 1970–79. *European Economic Review* 28:201–31.
- McGuckin, Robert. 1991. Multiple classification systems for economic data: Can a thousand flowers bloom: And should they? In *Proceedings of the 1991 International Conference on the Classification of Economic Activities*, 384–407. Washington, D.C.: U.S. Bureau of the Census.
- Millward, Neil. 1993. Establishment surveys in Britain: A boon to labour economists. In *Proceedings of the International Conference on Establishment Surveys*, 266–74. Alexandria, Va.: American Statistical Association.
- National Commission on Employment and Unemployment Statistics. 1979. *Counting the labor force*. Washington, D.C.: Government Printing Office, September.
- Orchard, Terry, and Peter Stibbard. 1993. Labour income surveys in the United Kingdom. In *Proceedings of the International Conference on Establishment Surveys*, 316–22. Alexandria, Va.: American Statistical Association.
- Pergamit, Michael R. 1987. The effects of unions on employment and hours movements in manufacturing establishments. Washington, D.C.: U.S. Department of Labor, Bureau of Labor Statistics. Unpublished paper.

- Rees, Albert. 1993. The role of fairness in wage determination. *Journal of Labor Economics* 11 (January): 243–352.
- Rosen, Sherwin. 1990. Comment. In *Fifty years of economic measurement*, ed. E. R. Berndt and J. E. Triplett, 295–98. Chicago: University of Chicago Press.
- Ruser, John W. 1993. Workers' compensation and the distribution of occupational injuries. *Journal of Human Resources* 28 (3): 593–617.
- Ryscavage, Paul M., and John E. Bregger. 1985. New household survey and the CPS: A look at labor force differences. *Monthly Labor Review* 108 (September): 3–12.
- Stafford, Frank. 1986. Forestalling the demise of empirical economics: The role of microdata in labor economics research. In *Handbook of labor economics*, ed. Orley C. Ashenfelter and Richard Layard, 1:387–423. New York: Elsevier.
- Triplett, Jack E. 1990. Reviving the federal statistical system: A view from within. *American Economic Review* 80 (May): 341–44.
- U.S. Department of Labor. Bureau of Labor Statistics. 1992. *BLS handbook of methods*. BLS Bulletin no. 2414. Washington, D.C.: U.S. Department of Labor, Bureau of Labor Statistics, September.

Comment Charles Brown

Marilyn Manser has given a very useful overview of the many data sets that can be used to study U.S. labor markets, an analysis of data use and main topics in papers appearing in major journals, and some indications of what will and might be done to improve the usefulness of the data. The section on data use in particular is valuable not only as a summary of what labor economists are doing but as framework for thinking about new data and better uses for existing data.

Finding little to complain about, I will instead hope to add a constructive outsider's perspective to three aspects of making the data more useful to researchers: adding content areas, generating matched employer-worker data, and increasing the use of establishment data.

Additional Content

Manser identifies a number of areas where the BLS has recognized a need for additional data collection efforts, either as "permanent" changes to regular programs or occasional studies. No discussant could resist the temptation to add a few more to the list:

Impacts of growing internationalization. While wage and employment data allow one to ask whether industries that export a significant share of output or have significant import penetration fare better or worse than other industries, much less is known at the firm or plant level. Have firms or plants that export

Charles Brown is professor of economics and research scientist at the Institute for Social Research, University of Michigan, and a research associate of the National Bureau of Economic Research.

a significant fraction of their output or import a significant fraction of their intermediate inputs done better than other plants in the same industry? How do wage and employment growth at firms with subsidiaries abroad compare with wholly domestic concerns?

Small businesses? Or new businesses? Despite the increased interest in job growth in small firms, regularly published data by firm size are scarce. And much of what we “know” may be misleading: the job growth attributed to small firms may in large part be growth of new businesses that happen to be small, rather than evidence of the robustness of small firms in general.

What goes on inside the firm? Much of the new theoretical work in labor economics has emphasized the imperfect information that employers have on both the ability of prospective workers and the productivity of those they hire. That has led to interest in how intensively firms collect information prior to hiring, how often less productive workers are discharged or encouraged to leave, the extent of pay variation within job grades (and the extent to which such variation is tied to performance rather than seniority), the importance of promotions in providing incentives, and whether voluntary turnover is concentrated among the most or least productive workers. The BLS has in the past successfully collected data on pay variation within job grades and the importance of performance in progressing through a grade. Significant improvements in our knowledge base here are possible by extending these efforts, even though some of these questions are best answered by an accumulation of firm-level case studies (see Prendergast, chap. 9 in this volume).

Areas dominated by “theory only” papers. Despite the admirable detail that Manser does present, it is difficult to judge from table 1.5 whether the “theory only” papers theorize about topics on which useful data might be collected, require data that realistically cannot be collected, or could (or have) been tested by other researchers with available data. It is worth asking whether a subset of the theory-only papers might help identify a few gaps that could usefully be filled.

Matched Employer-Worker Data

As Manser notes, matched data on workers and their employers has been on the wish list of labor economists for some time. One of the more promising recent data developments is the beginning of progress in assembling such data; indeed, three papers presented at the conference offer examples. The French data used by Abowd and Kramarz (chap. 10) match workers’ earnings records (from social insurance administrative files) to longitudinal surveys of their employers. Troske’s (chap. 11) data come from matching manufacturing workers’ decennial census records to their employers’ Census of Manufactures and Annual Survey of Manufactures data. Bronars and Famulari’s (chap. 13) data are

from a pilot project that used one-stop shopping, asking employers about detailed characteristics of their workers as well as about the establishments themselves.

Another approach is to begin with traditional household surveys, ask respondents to identify their employers, and then either contact the employer directly or locate employer data from an external file. The Health and Retirement Survey, which began in 1992 with in-person interviews of individuals born in 1931–41 and their spouses, obtained the names and addresses of respondents' employers in order to collect matched data on pension plans and health insurance coverage directly from the employers. Preliminary analysis suggests surprisingly high cooperation levels by respondents, success in locating employers, and employer cooperation rates similar to those obtained by the BLS in its surveys of employer benefit practices. Contacting only employers who offered pension plans or health insurance focused our employer surveys on relatively larger, more established employers.

We also have had some modestly encouraging experience with the employer lookup strategy. Starting with workers interviewed as part of a household survey, we asked respondents for employer name and address and attempted to find these employers in the Dun and Bradstreet files. Here we encountered some respondent reluctance, probably increased by the fact that we were obtaining the information by telephone, and some difficulty finding matching employer data, particularly when the address information was incomplete or the employer was a small business (Brown and Medoff 1996).

Taken together, the variety of strategies that have been used to produce matched data are one of the more optimistic developments for empirical labor economists.

Increasing the Use and Usefulness of Establishment Data

For me, the most striking single facts in table 1.4 are the increasing dominance of household survey microdata and the very limited use of data collected from establishments. I see three important reasons for the greater use of household data: household surveys have recognized the power of longitudinal analysis, they tend to have a broad, multipurpose focus and plentiful covariates for multivariate analysis, and they have found it possible to make microdata available to researchers. It is natural to ask whether establishment surveys might be made more useful in these three ways.

Longitudinal analysis. There are two barriers to matching across waves of an establishment survey to create longitudinal data. One is technical: rotation patterns that minimize compliance burdens for small establishments limit the number of periods that such establishments can appear. The other is traditional: because users of the establishment data have analyzed changes in employment, wages, and so forth, aggregated across establishments, the focus on establishment-level changes that would lead to routine creation of longitudinal files has

been missing. If other barriers to use of establishment data can be overcome, the traditional use will change quickly in directions that make longitudinal files irresistible; where the rotation pattern is a serious limitation, the possibility of stretching out participation (e.g., the CPS's four months in, eight months out, four months in rotation) will naturally arise once the usefulness of longitudinal analysis is established.

Broader focus. Compared to establishment surveys, household surveys often have fewer observations but more variables. This is particularly true when successive waves of a household survey that feature "one shot" supplements can be linked longitudinally, so that parts of different supplements can be analyzed together even though the data were obtained at different times. In principle, information about individual establishments can be broadened either by matching observations across surveys or by adopting the one-shot supplements as a regular feature of establishment survey programs. In practice, matching across surveys is constrained by institutional limitations that are uniquely American (so that, at present, the boundary that separates the BLS and the Census Bureau has not been overcome) and by the fact that small establishments are sampled with low probabilities in different surveys (so that, by design, few would have answered two different surveys that one might want to match). Still, the option of matching ES-202 data (which is available for virtually all establishments) for several years before *and after* a smaller survey is appealing and hopefully will become common practice. Supplements seem more promising. While there are concerns that longer surveys will jeopardize cooperation, short supplements to *several* waves of a survey can produce the same favorable benefit-burden ratio that one finds in household surveys. (One subtle cost of broadening an establishment survey is that a different "knowledgeable respondent" may be required for the additional information.)

Microdata. Concerns about confidentiality have nearly prevented nongovernment researchers from obtaining access to establishment microdata, and this in turn is largely responsible for labor economists' neglect of such data. A sensible policy on data release ought, however, to begin by recognizing the exceptions to the preceding sentence. Establishment data has, on occasion, been made available to outside researchers, often but not always with top-coding or other ways of blurring the identity of individual establishments. I probably have never gotten the *really* sensitive stuff, whatever that is, but that is my point—that "confidential" but not particularly sensitive data are excellent candidates for wider, if still restricted, availability. Last month's R&D spending is sensitive; employment—which is anyway available from Dun and Bradstreet's for most establishments and firms—is not.

Regional data centers that provide controlled access to such data for selected researchers are a welcome development. An alternative is to make the data available to selected outside users via contracts that specify prohibitive penal-

ties if the data are abused. The contractual approach is particularly helpful to researchers who cannot feasibly relocate to a data center; it has worked well for making detailed geographic identifiers available to users of household surveys.

For some research purposes, observations for individual establishments may not be needed but published tabulations represent an inadequate way of aggregating the data. More creative ways of grouping data can sometimes provide “almost micro” data (minimizing within-group variation) while preventing researchers from identifying individual establishments. For example, define cells according to a rather elaborate cross-tabulation (e.g., industry by region by *very* detailed size class) and then aggregate across adjoining size classes when case counts fall below some cutoff. Within-cell variances would come (almost) free; researchers could work with the means of the logarithms of variables rather than the logarithms of the means. As long as an (incremental) government employee could process one or two such requests per week, the costs would be comparable to one round-trip airfare to a data center.

Grounds for Optimism?

In addition to the explicit content of Manser’s paper, I detect an implied theme that is, potentially, very important. The historical mission of many federal surveys was to provide reliable aggregated data for a rather limited (if very important) set of variables. As long as that focus remained, innovations along the lines discussed in her paper and my comment were unlikely. Implicit in Manser’s paper is a broader view of what the establishment surveys are meant to do, and that is a critical first step in improving their usefulness, and so increasing their use.

Reference

Brown, Charles, and James Medoff. 1996. Employer characteristics and work environment. *Annales d’Économie et de Statistique*, no. 41/42.