

## STATISTICAL USES OF ADMINISTRATIVE RECORDS

David W. Cartwright, Bureau of Economic Analysis  
Paul A. Armknecht, Bureau of Labor Statistics

There has been considerable discussion in recent years of the notion that administrative records can and should be utilized more intensively to generate improved statistical series. Implicit in these references is the notion that administrative records constitute a new and largely untapped statistical resource.

The purpose of this paper is to demonstrate that administrative records are presently an important and integral part of the Federal statistical system. There will not be time this morning to present the entire paper [1]. I will cover some general characteristics of the use of administrative records, describe some of the uses at the Bureau of Economic Analysis and the Bureau of the Census, and suggest potential new uses of improved administrative record sets. Paul Armknecht will discuss a case study of the uses of unemployment insurance records.

### GENERAL CHARACTERISTICS

A major distinction between the use of administrative as opposed to statistical records is the degree to which the statistician is in control of the design and collection of the records. Statistical records and their collection procedures are designed, documented, and controlled to yield the desired statistical results. When administrative records are used, the statistician must locate the desired records, determine their suitability for the intended use, determine their accessibility, and locate documentation on the sources and methods used in collecting the records--often a formidable task. The statistician must also devise methods for overcoming technical problems inherent in the records.

Most statistical uses of administrative records have developed on an ad hoc basis. There are few examples of administrative record systems that have been designed with statistical uses in mind [2]. Only rarely have statisticians specified changes in the design or procedures of an administrative record system.

As Joe Knott [3] points out, statistical uses of the various administrative record sets have generally been uncoordinated. A body of uses and users have developed somewhat independently for each record set. There is very little standardization of the accessibility, documentation, format, and quantity of information available from the various record systems.

Statistical uses are often met with some resistance from the operating personnel of the collecting agency. This is partially due to the diffusion of responsibilities. Such statistical uses are often viewed by administrative personnel as an annoying addition to their already overburdened work schedules. Other reasons for this resistance are related to confidentiality

restrictions and the massive nature of the record sets. Even a seemingly minor change in the information to be collected or the collection procedures could have far reaching cost and timing repercussions.

Another characteristic of the use of administrative records is that they are used most heavily by those interested in regional statistics. Sample surveys provide a wealth of current information at the national level but it is not feasible to conduct regular surveys large enough to yield reliable local area data.

### CURRENT CENSUS AND BEA USES

Two agencies which make heavy use of administrative records are the Bureau of Economic Analysis (BEA) and the Bureau of the Census.

BEA relies heavily on the use of administrative records for its preparation of national economic accounts and related measures. In its role as both an intensive user and as a producer of government statistics, BEA has perhaps more experience than any other agency in the systematic use of a wide variety of administrative records. The problems of inconsistent definitions, uncoordinated formats and presentation techniques, and poor timing are familiar to the BEA analyst who must be alert to and make adjustments for deficiencies in the data.

The list of administrative record tabulations which are used directly to estimate components of the national income and product accounts, the input-output tables, and the international accounts is extensive and includes various types of tax records, regulatory records, financial records, customs reports, and budget documents.

BEA's regional accounts make particularly heavy use of administrative records since most periodic surveys are not designed to yield information at detailed geographic levels. For example, the primary source for estimates of wages and salaries at the county level is payroll data collected and tabulated by State bureaus of employment security as part of the Unemployment Insurance system. As with the national accounts, however, the regional accounts make use of a wide range of other administrative records including regulatory and financial records.

The Bureau of the Census is not only the primary collector of statistical information in the Nation but it is also a major user of administrative records. Administrative records are employed in a variety of ways by the Census Bureau, including: design and evaluation of censuses and surveys, identification of sampling universes, estimates for non-surveyed portions of the universe, and imputations for missing cells.

Census Bureau uses of administrative records in censuses and surveys are extremely important and have resulted in significant reductions in cost --and improvements in data quality. Most references to increased reliance on administrative records, however, refer not to these uses but to direct tabulation of administrative records in lieu of more frequent or more extensive censuses and surveys.

Census does utilize direct tabulations of administrative records to produce annual statistics for its County Business Pattern series. Census also, as Joe Knott noted [3], makes extensive use of administrative records in its program of current population estimates and money income estimates for the General Revenue Sharing program.

There are many other examples of Federal agency uses of administrative records. Not only are there many more direct uses, such as the use of Social Security Administration (SSA) records by health research agencies [4], but there are an enormous number of secondary uses, such as the increased use of BEA's regional income statistics for Federal fund distribution formulas.

#### POTENTIAL NEW USES

The problems of non-standardized concepts, definitions, and procedures among administrative record sets seriously inhibit statistical uses. The potential for major new uses of administrative records may, in fact, be quite limited because of these problems.

There are, however, important potential uses for improved administrative record sets. The record set which is currently being assembled by the Social Security Administration from IRS Forms W-2 and W-3, for example, illustrates this potential.

Employers must forward Forms W-2 for each person employed during the year to the SSA together with W-3 (establishment summary) forms. The units which employers currently use for establishment reports to the SSA are optional and differ widely among employers. If it were to become a program requirement for employers to use the establishment definitions and codes developed by the Census Bureau for its Standard Statistical Establishment List (SSEL), the resulting file of W-3 forms would be immensely useful for statistical purposes. If confidentiality problems could also be overcome, this procedure could actually bring about a reduction in paperwork burden for employers. By grouping W-2 forms into establishment units and providing a W-3 summary (pre-coded with Census Bureau establishment codes), the employer would provide an annual report of employment and wages by establishment for a variety of administrative and statistical purposes. If the SSEL could be used to code establishment by industry and geographic location (State, county, and possibly sub-county units), the resultant file would be useful to BEA for its measurement and analysis programs. It would also be useful to the Census Bureau for its County Business Patterns and other programs.

Future Continuous Work History Samples (CWHHS) will be drawn from W-2 forms. These longitudinal samples will continue to be useful for studies of worker migration and other types of labor force behavior. The inclusion of worker address information in the CWHHS would permit the construction of an annual series of commuting estimates. Current commuting data would be useful to the BEA in the construction of per capita income estimates for States and counties and to the Bureau of Labor Statistics (BLS) in the calculation of unemployment rates for local areas. Current commuting estimates would also be useful to transportation, energy, and housing planners.

#### THE UNEMPLOYMENT INSURANCE (UI) SYSTEM: A CASE STUDY

A case study of the statistical usefulness of administrative records for establishments can be gleaned from the unemployment insurance (UI) system. This system was established as part of the Social Security Act of 1933 to serve as a countercyclical income maintenance program for offsetting losses in wage and salary income of the experienced work force. Initially, UI covered only employers in the private nonfarm economy with eight or more employees. Over the years, the system has been continuously expanded. In March 1978, over 90 percent of employed workers were covered by the State and Federal UI system.

In the UI system, a variety of administrative data is maintained. Three important data sets which serve as the primary source of statistical uses are discussed in this paper (see Chart 1).

First of all, there is a master list of more than 4 million subject employers which contains the names and addresses of covered firms and both actuarial and statistical information. Secondly, information from the quarterly tax reports filed by employers is maintained. Finally, in all but 12 States, firms report the total wages paid to each employee during the quarter to determine an individual's eligibility and benefit amount when filing a UI claim.

Master List of Employers.--State agencies collect and process certain statistical information to help provide standardization for reports and tabulations. Employers are assigned county and industry codes. Industrial activity is reviewed on a three-year cycle, and attempts are made at identifying multi-establishment employers and setting in place a mechanism for supplemental reports of employment and wages by county and industry. The UI list is used by State agencies to draw samples in the Federal-State programs sponsored by BLS and operated by the States. A number of States also use the list to publish industrial directories. The lists are provided to the Bureau of Labor Statistics to use for sampling purposes under a pledge of confidentiality. BLS uses the lists to develop its UI Name and Address File which serves as a sampling frame for its directly collected surveys.

The UI Name and Address File has a number of drawbacks. Since it is derived from an administrative source, many of the refinements needed for sampling purposes are not present. For example, the major identifying field in the file is a UI account number which is assigned independently by the various States. There is no unique way to identify firms or companies within a corporate structure across States. Also, identification of multi-establishment employers varies from State to State.

Employers' Quarterly Tax Report.--Taxes are collected quarterly from subject employers by mailing each employer a tax form on which he reports the total wages paid to employees during the quarter, the amount of these wages that is subject to taxes, the taxes due, and the number of employees on the payrolls for the period that includes the twelfth of each month. The tax forms are due at the State agency 30 days after the end of the reference quarter. Multi-establishment employers are also mailed a statistical supplement with their tax report requesting a breakdown of the monthly employment and wage figures by reporting unit. Five months after the end of the quarter, State summaries in machine readable form are sent to BLS, Washington. Two summaries are required of each State: (1) Statewide by four-digit industry, and (2) counties by two-digit industry. States that can provide four-digit industry by county, need only send one summary. These summaries are called ES-202 reports.

Many programs of the BLS and BEA rely on the ES-202 report's employment and wage data. Within BLS, the Current Employment Statistics (CES), Labor Turnover Statistics (LTS), the Occupational Employment Statistics (OES), Industry Projections, and Occupational Safety and Health Statistics (OSHS) programs are benchmarked to industrial employment data emanating from the ES-202 report. The BEA National income and personal income estimates rely heavily on the UI administrative data. In addition, personal income is used in formulas to allocate billions in Federal funds to State and local governments. At the local level the average wages of workers covered by UI are used to adjust the average annual wage payments allowed Comprehensive Employment and Training Act (CETA) Public Service Employees. The State agencies also make substantial use of the employment and wage data to assess the economic vitality of local labor markets in their labor market information programs. Practically every employment-related statistic that is generated in the BLS-BEA-State employment agency enclave has the UI administrative data as its base.

The ES-202 report has its limitations and problems. There is no set mechanism of quality control to assure that all subject employers are reporting. There is no program of quality assurance for ascertaining the accuracy of data reported by employers on their tax reports. Statistical reports which are a by-product of an administrative program often receive a low priority. The statistical functions in producing the ES-202 report compete for basic UI program re-

sources with tax collection, benefit payment, and research activities. Hence, many States cannot fully implement industry coding and multi-establishment "breakout" activities.

Individual Wage Records.--In most States, the collection of the quarterly tax reports also involves an itemization of individual workers' wage payments identified by social security number. This data base provides a rich source of information on an individual's earnings history. The Current Wage and Benefit History program of the U.S. Department of Labor is attempting to tap this data base to link earnings experience with workers' eligibility and receipt of UI benefits. Since each individual's earnings are linked to the employer, studies on wage dispersions by industry and county (on a place of work basis) are feasible. These files are also being used to map mobility patterns and labor turnover actions as part of Labor's Employment Service Potential program.

Improving Data Quality.--The UI administrative data have room for improvement because of the large and cumbersome task of identifying multi-establishment employers. Their major strength is the quarterly collection and timeliness versus other sources of establishment records--namely, the Census Bureau's County Business Patterns program. Census does considerable work annually in identifying and maintaining multi-establishment breakdowns of firms in its Company Organization Survey. Access to these data could help identify and refine multi-establishment reporting problems in the UI record system.

At the same time, one of the weaknesses of the Census' establishment records is the industry codes of single-establishment firms. Those single unit firms not covered in the 1972 or 1977 Economic Censuses retain industry codes assigned from information submitted when the application for an EI number was made. A matching of industry codes in the two data systems could improve the coding of single establishment firms on the Standard Statistical Establishment List and help identify potential problem areas between the two systems; i.e., such a match could determine how much of the difference between BLS and Census series is due to coding, how much is due to reporting differences, and how much is the result of differences in treatment by central administrative offices.

#### CONCLUSION

Administrative records are well integrated into the Federal statistical system. Their uses are numerous, complex, and diverse. Statistical uses, as well as the basic record files available for statistical use, have developed somewhat independently of one another. Problems with accuracy, consistency, and confidentiality of existing administrative records seriously limit potential statistical application. There is, however, enormous potential for improved and better coordinated administrative records to provide much-needed statistics on a current

and geographically detailed basis. These potential uses could yield improved statistics, at lower cost, with less reporting burden on the public.

NOTES AND REFERENCES

- [1] The full paper can be obtained by writing to Dave Cartwright, Chief, Regional Economics Information Systems Branch, Bureau of Economic Analysis, BE-55, 1401 K Street, N.W., Washington, D.C. 20230.
- [2] An exception to this might be the treatment of race on social security records. See: Alexander, Lois, Statistical Program of Administrative Records: Some Legal Issues,
- [3] Knott, Joseph, Major Administrative Record Files, American Statistical Association 1979 Proceedings, Section on Survey Research Methods.
- [4] There was an entire session devoted to this topic at the 1979 meetings entitled, "Linked Administrative Statistical Samples for Mortality Research." See also: Hirschberg, David A. and Vernon Renshaw, Access to Administrative Records on Establishments and Individuals for Public Policy Analysis, American Statistical Association 1979 Proceedings, Section on Survey Research Methods.

Chart 1. Statistical Uses of Unemployment Insurance Administrative Records from Establishments

