MATCHING ECONOMIC DATA TO THE SURVEY OF INCOME AND PROGRAM PARTICIPATION: A PILOT STUDY

Sheldon E. Haber, The George Washington University
Paul M. Ryscavage, Douglas K. Sater, Victor M. Valdisera, U.S. Bureau of the Census

The new Survey of Income and Program Participation (SIPP) will undoubtedly become a major source of data on a wide variety of aspects of the well-being of our nation's households, families, and individuals. The very richness of SIPP suggests the desirability of augmenting it with micro-level establishment and enterprise data from the economic censuses and other data files maintained by the Bureau of the Census, since the marginal cost of merging these data with SIPP is relatively small and the potential gain in knowledge is very large. One area where the payoff relative to cost of enhancing SIPP is sure to be substantial and significant is that pertaining to the behavior of labor markets.

A list of some of the areas in which a SIPP-economic data file can yield new insights includes the following topics:

- The relationship between capital and wage rates
- Labor mobility
- Low wage workers and low wage firms
- Measuring the effects of minimum wage legislation
- Structural unemployment
- Identifying high tech workers and high tech firms
- Implications of the transition from a goods to a service economy
- Unions and the labor market
- The substitutability of capital and labor
- Productivity analysis

Besides the substantive knowledge to be gained by merging SIPP demographic and economic data, there are externalities associated with merging these data sets. First, it will be possible to verify the accuracy of the size of firm estimates given by respondents in survey data. An additional, indirect benefit of linking SIPP and economic data stems from the fact that the former is a representative sample of the working population. Matching on work place will yield a stratified sample of firms where the probability of selection is inversely proportional to firm size. By weighting the number of firms in each size group, estimates for the entire population of firms can be derived. The sample of employers would be contained in a single data set versus the diversity of data sets in which the economic data are now found, with the same format across employers. These advantages plus the manageable size of the sample should provide valuable insights into the structure of production within and across sectors of the economy at a point of time and over time.

1. SIPP and the Economic Data Files In merging demographic and economic data, it is necessary to know the information contained in the various files to be linked and how each file is constructed. In this section, we briefly describe four data sets which might be incorporated into a SIPP-economic data file.

SIPP contains demographic and program related data. Economic data are found in the Standard Statistical Establishment List (SSEL), the Longitudinal Establishment Data (LED) file, and the enterprise statistics (ES). The SSEL covers all establishments and companies with employees and yields current information on employment and payroll. The LED, as its name implies, contains longitudinal data but is restricted to manufacturing establishments. The ES, on the other hand, covers companies in the construction, mineral, manufacturing, wholesale trade, and retail trade industries, and most service industries.

The SSEL is a complete directory of establishments in single and multi-establishment enterprises with one or more employees, irrespective of industry. The SSEL links parent companies, subsidiaries, and their establishments. It contains information on approximately 4.7 million enterprises and 5.7 million establishments.

The importance of the SSEL is that it is a current file containing a complete list of establishments and companies with paid employees. While the SSEL contains a narrow range of economic data, these data impart valuable information. For example, the SSEL contains the address of the physical location of establishments which is useful for merging the demographic and economic data, since it is a primary link in identifying an individual's place of work. Employment and payroll figures yield an estimate of average annual earnings, thereby indicating whether an employer is a low or high wage employer.[1] Sales and employment figures provide a proxy measure of productivity. Operational status information can be utilized to identify establishments which have become inactive. Additionally, the SSEL contains longitudinal information. Currently, establishment and company data are carried for two years in the SSEL.

The LED is a longitudinal micro-data base containing data at the establishment level from the Annual Survey of Manufactures and the Census of Manufactures.

The LED provides a much broader range of information about establishments than the SSEL. For each manufacturing establishment, value added per production worker, which is a measure of labor productivity, can be calculated. For the larger establishments with 250 or more workers, information is available on depreciable assets and rented machinery so that capital/labor ratios can be computed. Also, a better measure of labor compensation, including fringe benefits, can be obtained.

Like the Census of Manufacturers, the enterprise statistics (ES) are collected

every five years. These data cover enterprises whose primary activity is in an in-scope industry. For each enterprise, the data are consolidated over all operating units. The information contained in the ES is similar to that in the Census of Manufactures, except that fringe benefit, asset, and related data are available only for companies with 500 or more workers.

2. Some Applications of Micro-Demographic and Economic Data In this section, three applications of a SIPP-economic data file are discussed to illustrate how this data set can help explain policy issued relating to earnings and employment.

A. Low Wage Workers and Low Wage Firms While survey data such as the CPS provide insights into the characteristics of low wage workers, they provide no information about low wage firms. Despite this lack of information, an a priori description of low and high wage firms can be formulated. All else being the same, low wage firms will be labor intensive and, hence, tend to be smaller than high wage firms. And because recruitment and hiring cost relative to the level of wages will tend to be high, low wage firms will also advertise less for labor and employ fewer screening devices to weed out suitable workers; thus, their work force will be of lesser quality than their high wage counterparts. Less qualified workers, on the other hand, e.g., younger workers and those who are less educated, will be attracted to low wage firms because their marginal product is less than that required to gain employment in high wage firms. More generally, workers with given characteristics and tastes sort themselves among firms with similar requirements for labor.

Corresponding to the greater prevalence of low quality workers in low wage firms, one might expect that in these firms (vis-a-vis high wage firms) a higher proportion of capital expenditures is for used rather than new machinery and equipment; likewise, the proportion of depreciable assets retired each year is likely to be smaller in such firms. Furthermore, given that labor is of lesser quality and capital is of an older vintage, it would not be surprising if value added per worker were relatively low in low wage firms.

Other characteristics are more easily seen by focusing on high wage firms. To the extent that high wage firms are capital intensive, their need for trained workers is likely to be greater than that of low wage firms. Capital intensiveness suggests greater use of resources to monitor output; hence, a higher proportion of the work force may be needed in supervisory positions. To reduce turnover, which disrupts the production process, high wage firms will substitute future benefits in the form of pensions for current benefits in the form of wages. A SIPP-economic data file would permit verification of these hypotheses.

Information about low and high paying firms is important for another reason besides the light it sheds on how production is organized in these two types of firms. Since low paying firms are a source of employment for workers with relatively low productivity, it is of some interest to inquire into the extent to which low pay among workers is attributable to their employment in such firms. In approaching the question of why some workers are paid less than others in this manner, low wage employers can be viewed as providing employment opportunities with attendant low earnings, not because they discriminate against certain groups of individuals, but because the production processes that are most efficient for their mode of operation do not require high quality labor and, furthermore, they inhibit payment of high wages.

A procedure for verifying this view would be to sector firms according to whether they are low paying or high paying. With this sectoring of firms, one would expect, as indicated above, that the mix of workers and capital is dissimilar between the two sectors. Assuming this is so, to what extent are differences in individual earnings in low and high paying firms due to the characteristics of the workers and capital employed in each type of firm? Also, to what extent are workers with similar characteristics renumerated in the same way in each type of firm? The answers to these questions can be obtained from a SIPP-economic data file.

B. Structural Unemployment An issue of long standing is what happens to workers who are displaced from their job as a result of structural disequilibria. How long do they remain unemployed vis-a-vis other workers who separate from an employer? What sources of income, including cash and noncash government transfers, do they draw on when they are unable to find work? When they find a job, how do earnings in the new job compare to earnings in the old one? If there is an earnings loss, how much of this loss is recouped, say, after two years?

A major problem in answering these questions is that workers do not know if they are structurally unemployed. One way of identifying such workers is to ascertain what has happened to the firms in which they were last employed.

If the firm has undergone a substantial decline in employment or has closed down for a a relatively long period of time, say, longer than the typical recession, one may presume that it has undergone a shock which is typical of the shocks experienced by firms subject to structural disequilibria. It also can be presumed that the employees of these firms experience the aftereffects of such shocks.

A SIPP-economic data file would enable one to determine the extent to which firms are subject to severe, long-term shocks as evidenced by plant closures and substantial reductions in employment, and how such shocks affect their work force.

C. High Tech Workers and High Tech Firms Despite the importance of new technologies for improving productivity, there is no widely accepted definition of a high tech industry. Based on a definition which includes industries with a ratio of technology-oriented workers[2] to all workers of at least 1.5 times the industry-wide average, Riche, Hecker, and Burgan (1983) estimate that 13.4 percent of all wage and salary workers were employed in

high tech industries in 1982.

High tech industries have been cited as having a large group of high and low wage workers whereas other industries are comprised of workers who are concentrated in the middle of the earnings distribution. While it is useful to know how workers in high tech and other industries differ and the differential growth of employment in the two kinds of industries, it is equally important to know the characteristics which differentiate high tech from other firms and the differential in the rate of growth of the two types of firms.

As is self-evident, not all firms in high tech industries utilize the latest technology, and new techniques of production are utilized by firms in industries besides those labeled as high tech. One approach to distinguishing between the two types of firms would be to compare the characteristics of the industries denoted on a priori grounds as high tech with other industries and then to use this information to identify high tech firms. To illustrate this approach, assume that the a priori criterion used to denote high tech industries is the one noted above, namely, that the ratio of high tech to all workers in a given industry to the similar ratio for all industries is higher than some minimum value. Assume also that the high tech industries exhibit high values of the following ratios: capital expenditures for new computers to all capital expenditures, capital expenditures to asset value, and capital to labor. Given a set of characteristics which permit the bifurcation of industries, the multivariate technique of cluster analysis can then be applied to identify high tech firms within both high tech and other industries.

The outcome of the cluster analysis is a partitioning of firms into categories, i.e., high tech and nonhigh tech firms, as determined by the data, where each cluster of firms represents a homogeneous set of observations. An advantage of applying the aforementioned two-stage procedure using a SIPP-economic data file is that it provides an independent test of how well the procedure works. For if the approach is successful, the proportion of workers who are technology-oriented among the firms classified as high tech (taken as a group) will be higher than the similar proportion for firms classified as nonhigh tech (again, taken as a group), and the difference in proportions will be greater than the corresponding difference when industries are classified as high tech and nonhigh tech.

Having identified high tech firms, in contrast to high tech industries, insights can then be obtained as to how production processes in these firms differ from their nonhigh tech counterparts. At the same time, it will enable one to better define high tech occupations and how workers in these (and other) occupations in high tech firms differ from similar workers in nonhigh tech firms.

3. The Pilot Study A principal part of the pilot study is designed to assess the availability, sources, coverage, and content of the various economic data files maintained by the Bureau of the Census and to explore study areas and issues to which a data set combining micro-worker and firm data would be applied. In the course of this study, specific demographic and economic variables have been identified which should be incorporated into such a data set. Additionally, it was anticipated that methodological problems inherent in this undertaking would be revealed; indeed, this has been the case.

A second phase of the pilot study is to investigate the efficiency of four alternative methods of identifying an individual's employer. Each method is based on different information for searching the SSEL and identifying the employer's census file number (CFN). The first utilizes information on employer name, the state of residence and/or zip code of the employee, and census industry code. The same information is used in the second method; however, additional reference materials, e.g., 1980 Census Company Name and Place of Work lists, Dun and Bradstreet reference books, Standard and Poor directories, and telephone books, will be used to obtain the exact address of an individual's employer. The third method uses the employer's name and exact address if known. In the last method, if the employer's identification number (EIN) is known, it is used in conjunction with the information available in the first three methods to identify the employer's CFN. For each method, match rates and cost information will be developed for a small sample of workers.

A third phase of the study is the construction of a pilot SIPP-economic data file in which the SIPP portion of the file would be restricted to full-time workers in large manufacturing establishments; the source of the economic data would be the LED. The objective in this phase is to calculate match rates between workers in SIPP and their establishments in the LED.

Given the importance of the wage determination process, one of the areas noted above, e.g., low wage workers and low wage firms, would be studied when the pilot work file is completed. Demonstration of the utility of this research endeavor in terms of its contribution to the economic literature would constitute the final phase of the pilot study.

4. Methodological Problems in Matching Demographic and Economic Data In this section, attention is focused on two methodological problems. The first problem deals with procedures for tying workers to their establishment and company. The second relates to the estimation of data, in particular, asset and fringe benefit data, which although available for large establishments and companies, are generally not collected for small ones.

Central to the creation of a SIPP-economic data file is the ability to determine the establishment and/or company in which a person is employed. The most promising and least expensive way of doing this is to match on firm name and physical address of an individual's place of work. This information will be available in SIPP and is available in the SSEL. Although the physical address is not necessary for identification of an individual's work place, its availability greatly facilitates

such identification since a firm may have more than one establishment in a local area.

For employers with only one establishment in an area, the firm name and employee's address will typically be sufficient to determine where a person is employed. As noted, for companies with more than one establishment in an area, the firm name and employer's address should be sufficient to identify the place of work. If an employer has more than one establishment in an area and the place of work cannot be determined using the employer's physical address, or no address is available, other information in SIPP can be utilized. Firm name, respondent's address, census industry code, and respondent's estimate of size of establishment and company can be used to identify a person's work place. For example, it is unlikely that a firm manufacturing bottles will have more than one large plant in a local area.

Another aid in identifying an individual's work place is the EIN. While a company may have a number of establishments in a local area, its subsidiaries, when identified by their own EIN, may have only one establishment in the area. Thus, the EIN of the employer for whom an individual works can be sufficient to uniquely determine the establishment in which that person is employed.

In the event that an unique work place cannot be determined for a multi-establishment firm, the employer's characteristics can be imputed. Data from the SSEL on number of employees and payroll can be averaged over a company's establishments in a local area. From the ES file, average values can be computed for variables not contained in the SSEL. For example, the average capital/labor ratio for a company with a chain of fast-food stores can be used as an estimate of the capita/labor ratio for each store in the chain.

Where it is not possible to identify a worker's firm by name in the SSEL, imputations can be made by averaging over establishments in the same local area and with the same census industry code as that of the given employer. Additionally, it may be possible to refine the imputation process by considering information contained in SIPP, e.g., the size of establishment in which an individual works and whether the firm has one or more than one establishment.

As indicated, information on assets and fringe benefits is not generally available for small establishments, but such information is available for a large sample of small establishments in manufacturing. Despite the fact that asset information is not collected for many of the firms in which individuals work, the use of an economic model, including industry, firm size, and other variables, may enable one to obtain reasonably accurate estimates of capital for small establishments.

Economic theory sugests a number of relationships which influence the amount of capital that a firm employs in its production process. In particular, since capital intensity varies with establishment size in closely related industries, it seems reasonable to assume that information about the number of employees in an establishment can be used to further refine estimates of its capital assets. All else being the same, one would expect the smaller an establishment, the lower would be its capital/labor ratio. [3] Additionally, holding everything else constant, including establishment size, low wage establishments will substitute labor for capital in order to economize on the use of the relatively expensive factor, i.e., capital. Thus, low wage establishments will tend to have a lower capital/labor ratio than high wage establishments.

Even among establishments of the same size whose wage rate is also the same, one would expect a lower capital/labor ratio the higher the proportion of production workers among all workers. When the proportion of production workers among all workers is high, or conversely, when the percentage of workers who supervise production is low, this comes about because a firm has few assets, relative to labor, to monitor. Additional relationships between assets and other variables may exist. For example, it may be that newer establishments in an industry are more capital intensive than older ones; likewise, regional variations in entrepreneurial ability may give rise to corresponding variations in capital intensity.

Besides economic relationships, engineering relationships also may be useful in estimating capital intensity. For example, it is plausible that an establishment's capital/labor ratio is positively related to purchased electricity per employee.

Finally, an economic model can also be utilized to estimate fringe benefits for small establishments and small companies. It is plausible to assume that fringe benefits in a firm are related to its size, average wage level, legal form of organization, industry, and region where it is located. With a SIPP-economic data file more refined estimates of fringe benefits per employee can then be obtained by taking account of the percentage of employees who are covered by life and medical insurance and a private pension plan in a given group of firms, say, (small) high paying establishments in manufacturing. Given this information, the average value of these benefits per covered and noncovered worker can be calculated for establishments in the group. Such information could provide a basis for imputing an important component of private non-cash benefits to individual workers. Although it should be evident from the discussion of this paper, this last illustration is indicative of the benefits to be derived from a SIPP-economic data file.

Footnotes

[1] The data referenced in this section as well as the remainder of the paper are available in the economic and (where appplicable) in the SIPP data collected by the Bureau of the Census.

[2] These are defined as engineers, life and physical scientists, mathematical scientists, engineering and science technicians, and computer specialists.

[3] An estimate of a firm's assets can then be obtained by multiplying the capital/labor ratio estimate by the number of workers in its employ.

Reference

Riche, Richard; Hecker, Daniel; and Burgan, John, "High Technology Today and Tomorrow: A Small Slice of the Employment Pie," Monthly Labor Review, November 1983, pp. 50-58.