Labor Market Analysis Using SIPP

Gary S. Fields and George H. Jakubson

Cornell University

Abstract

This paper examines the potentiality of the Survey of Income and Program Participation (SIPP) for labor market analysis. We consider five areas of analysis: (1) labor force participation, employment, and unemployment; (2) labor market effect£ of income maintenance programs; (3) earnings; (4) work and retirement of the elderly; and (5) migration. We find that the SIPP is a potentially rich resource for labor market analysis, surpassing much of what is to be found in existing databases. We note some remaining problems and make recommendations for changes.

INTRODUCTION

This paper examines the potentiality of the Survey of Income and Program Participation (SIPP) for labor market analysis. Six characteristics of SIPP render it a potentially rich resource for labor market analysis, surpassing what is to be found in existing databases:

(1) Its *Longitudinal Nature*. SIPP is longitudinal. This contrasts with other widely used data sets such as the CPS. Though these other data sets have the advantages of large sample size and familiarity, the cross-section nature of the data presents serious difficulties for the analyst, for two major reasons: First, labor market behavior is more appropriately viewed as a dynamic process. Second, we can never measure all of the relevant characteristics of an individual, and while methods exist to deal with these omitted variables' problems, they typically require multiple observations on individuals (Joreskog and Goldberger, 1975; Jakubson, 1983).

(2) *Sample Size*. SIPP will be two-thirds as large as a CPS, yet will be a panel study. The problems of insufficient sample sizes for population subgroups which arise in other panel data sets such as the Panel Study of Income Dynamics (PSID), the National Longitudinal Study (NLS), and the Retirement History Survey (RHS) will thereby be averted.

(3) *Continuity of Measurement*. For each sample panel, SIPP interviews will be conducted every four months for eight waves over 2i years. Information will be gathered for each month in the preceding four. This permits more continuous monitoring and diminishes a problem faced in virtually all other data bases: truncation of spells of unemployment and other labor market variables by calendar year[1]

(4) *Care of Measurement*. As a survey specifically designed to measure income from various sources, SIPP contains more comprehensive questions and cross-checking of labor

market earnings and nonlabor income than is typical. The links between labor and nonlabor income can also be carefully explored.

(5) *Matching with Other Data Sources*. It is planned to match SIPP with personal income tax records, Social Security earnings records, beneficiary records, and establishment records. It is critical that the proposed matching of SIPP to other data sources be completed in order to fully realize the potential of the SIPP data. Some possible uses of these matched data are outlined below.

(6) *Checking Possible Recall Error*. The staggered interviewing schedule will allow us to investigate how the length of time over which recall is made affects the accuracy of the answers. This is important not only for analyzing the SIPP data but for using other existing databases as well.

The balance of this paper discusses the potentiality of SIPP for analyzing the following aspects of labor markets:

labor force participation, employment, and

unemployment;

labor market effects of income maintenance

programs;

earnings;

work and retirement behavior of the elderly;

migration.

LABOR FORCE PARTICIPATION, EMPLOYMENT, AND UNEMPLOYMENT

The process generating spells of employment, unemployment, and labor force participation is a continuous one. The CPS and other such surveys measure labor force status at the time of the survey. This leads to well-known problems in estimating the duration of spells of unemployment (Chamberlain, 1982; Heckman and Singer, 1984). We know only the duration of unemployment up to the time of the survey but not the length of a completed spell of unemployment The March supplement to the CPS offers only limited help. It measures total weeks of unemployment in the previous calendar year. However, we cannot distinguish between a single spell of unemployment lasting 12 weeks and three spells, each of which lasted four weeks. The PSID, until the most recent waves, collected the same information, though over a number of years. Another difficulty with existing data sets is that we cannot measure spells of unemployment that cross over calendar years. For instance, a spell of unemployment from December 1, 1980 until March 31, 1981 would appear as four weeks of unemployment in 1980 and 13 in 1981, while the duration of the spell was in fact 17 weeks spanning two calendar years.

In view of these problems, job search and job matching models have been estimated on more limited data sets. The two most commonly used for this purpose are the control group from the SIME-DIME experiments and UI administrative records. The SIME-DIME data, used by Kiefer and Neumann (1979), Flinn and Heckman (1982), and others suffers from a lack of representativeness. UI administrative records, used by some authors, e.g., Katz (1977), also suffer from nonrepresentativeness, because the insured unemployed now represent fewer than half of all unemployed, and also from an additional problem: the data on the duration of

unemployment are usually truncated at the point at which benefits are exhausted. Still, these data sets offer better unemployment data than does the CPS.

SIPP offers the opportunity to reconstruct the true history of the process for a large representative sample of the population. But for this opportunity to be realized, it is critical that the calendar in the questionnaire be reproduced on the data tapes and not aggregated, so that researchers can see the true history of the process under investigation. It is planned to include weekly data in the SIPP public-use sample. This would be excellent.

A weakness in the core questionnaires of the SIPP data is that it distinguishes layoffs but not other causes of unemployment or nonparticipation. For both research and policy purposes, it is important to go further. We would want to distinguish, for example, new entrants to the labor force from those who are unemployed because of a plant closing. This information is available on a limited basis in Wave 5, Part 3.C.

One would also want to know how many hours were spent searching in each week of unemployment. A question or two ascertaining this information would be desirable. The information on hours of work in SIPP is unfortunately not as good as the information on labor force participation and unemployment duration. There is a CPS-type question about usual hours per week, but it is impossible to determine whether there were fluctuations in hours over the employment spell. This is particularly important if one wants to analyze underemployment and overtime. It would be helpful to include a question about whether the usual hours were worked during each week worked, and if not, then how many were worked and why it was different.

The SIPP can also be used to analyze mobility between jobs. Since monthly turnover rates in manufacturing are around 3%, we should see a substantial number of job-to-job transitions in the data.

In sum, SIPP may permit valuable new insights into processes of labor force participation, employment, and unemployment. To maximize its usefulness, presently planned variables will have to be coded carefully and additional variables should be collected.

LABOR MARKET EFFECTS OF INCOME MAINTENANCE PROGRAMS

SIPP is remarkable for the quality of the data on program participation. SIPP will go beyond existing data sets in containing information on

1. Whether respondents were recipients of any income maintenance program;

2. In Which programs, both cash and noncash, they participated;

3. How much they received from each;

4. Month-to-month reports on each of the above.

For the first time in a nonexperimental setting with large samples, it will be possible to accurately construct the budget constraint facing program participants and nonparticipants. This is the key to estimating the labor market consequences of income maintenance programs.

Most other data sets provide annual figures for welfare receipts, earnings, hours, etc. However, in most income maintenance programs, the relevant accounting period is a month, not a year. Owing to lack of data, previous attempts to carefully model the budget constraint (e.g.,

Hausman, 1981; and Moffitt and Nicholson, 1982) were hampered by the lack of monthly data, which required them in effect to assume that behavior is the same over each of the 12 months of the year. This is not true in general.

Knowing when individuals participated in income maintenance programs is essential for estimating their labor market effects. However, some needed information, especially on work expenses and child care expenses, is lacking in the core questionnaire. A topical module addresses these concerns, but it is only administered once to a panel—not often enough to be of much help.

Incomes from income maintenance programs are notoriously underreported. Assuming that the SIPP data will be matched with the administrative records from transfer programs, the amount of underreporting can be estimated. Underreporting takes two forms: (i) understating of receipts, and (ii) nonreporting of receipts. Which of these two is most prevalent has important implications for studying the dynamics of program participation, using both SIPP and other data sources. Of course, some legal protection for the recipients is necessary, much as is done with the quality control surveys for AFDC and other programs.

EARNINGS

The earnings information in SIPP is similar to that in the CPS, but not exceptional. The SIPP survey elicits the regular rate of pay, as well as gross earnings, from up to two employers for each four-month period. In addition, the yearly wrap-up, which includes information from the income tax returns, should enable one to calculate an average net wage for the year.

SIPP does not distinguish such things as overtime premia (in the same way that it did not distinguish overtime hours). Therefore, if reported earnings exceed the usual hourly wage times the usual weekly hours times weeks worked, the discrepancy cannot be accounted for. There is no way to distinguish a wage change from extra hours of work, and no way to determine whether the extra work hours receive higher pay. These could be corrected without substantially increasing the length of the survey.

It would also be interesting to ask the subjective question "If you were to work an additional hour at your regular wage, how would your total income after taxes and transfers be affected?" Although subjective, this would help assess people's perceptions of tax and transfer programs.

SIPP offers substantial information for explaining earnings differentials. The standard human capital type variables which typically enter wage and earnings equations are years of schooling, experience, union status, and tenure on the job. SIPP includes these variables as well as such others as field of study, actual experience with present employer, and time in the occupation that are potentially rich for earnings function analysis.

WORK AND RETIREMENT OF THE ELDERLY

Two data sets comprise the basis for most of the research done on work and retirement decisions of the elderly: the LRHS and the NLS. Both are panel studies. Both are rich in information on the personal characteristics of the respondents. Both contain measures of the Social Security benefits available upon retirement; in the LRHS but not in the NLS Social Security benefits are computable from the individual's Social Security earnings record, which is

appended to the file. However, both are extremely weak on private pension variables, since in neither data set was the structure of pension benefits obtained from the employer.

Because of the absence of pension information, some researchers have estimated retirement models using data on workers in one or a small number of firms whose pension rules were known; see Burkhauser (1979), Burtless and Hausman (1982), Fields and Mitchell (1984). Large differences were found between actual pension benefits and those estimated from industry or occupational data.

At present, therefore, investigators of the labor market behavior of the elderly must choose between representative coverage or accurate retirement income information; no existing data set offers both. SIPP presents a possibility of remedying this deficiency.

SIPP asks various questions about the nature of retirement benefits but does not determine the amounts of such benefits. The SIPP data will be matched with Social Security earnings histories, so Social Security benefits for alternate retirement ages can be accurately assessed. But in addition, because SIPP can also be matched to establishment data, this gives the opportunity of finding out from employers of older workers what their pension benefits, if any, will be upon retirement. (See this issue, Haber.) Although it is not now planned to do so, this would best be done by asking employers what pension worker X would be eligible for were s/he to retire at each of several ages (60, 62, or 65).[2]

With this additional information, SIPP would permit empirical research into a question hitherto unanswerable with reliable data: for a representative sample of older workers, what role does the structure of pension benefits play in their retirement decisions? In light of the importance of retirement research for public policy, this is a key question to answer.

MIGRATION

Studies of labor force migration are of two basic types:

1. *Migration Patterns*. Some studies are concerned with geography—where people move

    and where they come from. Among the explanatory variables used in such analyses are

    income opportunities, job prospects, and other labor market and nonlabor market data

    (availability and generosity of AFDC benefits, for example). Tabulations and Public Use

    Samples from the Population Censuses have been the predominant databases for this type

    of research.

2. *Who Moves When*. Migration is associated with such key life course/labor market events

    as attending college or university, labor force entry, career moves, and retirement.

    Longitudinal data sets contain information on these important life events and thus have

    increasingly been used for this purpose.

What has been lacking until now is information on a sufficiently large sample to offer

enough instances of such relatively rare events as retirements while at the same time containing;

adequate information on labor market conditions in the origin and in alternative destinations.

SIPP affords the opportunity to remedy this lack. But to do so, two important conditions must be

met: (i) Planned matches with other data sets must occur, and (ii) the geographic information

must be coded sufficiently finely. Matches with other data sets are important to assess accurately

the personal situation of the actual or potential migrant. A fine coding of geographic information,

perhaps combining identifiers of the largest SMSA's with status, is needed so that relevant labor

market and other data can be included by individual researchers, since it may not be obvious Which are the relevant variables to include. for instance, past research by Fields (1976) demonstrated that employment opportunities are instrumental in explaining migration flows but that the local unemployment rate is a very poor measure for this purpose; rates of new hires, quits, and layoffs perform far better.

For more on the potentiality of SIPP for migration research, see Dahmann (1984) and Comments by David (1984).

MISCELLANY

SIPP has a rather short period, $2\frac{1}{2}$ years, over which individuals are followed. A longer follow-up period would greatly enhance the feasibility of using SIPP to study dynamic aspects of the labor market It is important to attempt to distinguish individual decisions from overall cyclical variations. A 2i-year follow-up period does not seem long enough to do that. In addition, even though the sample sizes are relatively large, the short follow-up period may make it difficult to study relatively rare events (e.g., marital disruptions, plant closings) and next to impossible to study the aftereffects of such events (e.g., what happens to the wage rate of a worker whose plant closed?). There is a proposal to follow up only those individuals who have experienced particular events, for instance, marital disruption. (See this issue, David.) This is desirable only if budgetary limitations preclude longer follow-ups for the full sample. An alternative compromise would be to follow up a randomly selected subsample.

A few final points:

1. It is important for research purposes that imputted data continue to be marked as such.

2. As planned, detailed descriptions of the procedures used to derive the sampling weights should be made available as part of the SIPP user's guide.

3. Although the sampling frame does not include the institutionalized population, it would be useful to distinguish between attrition and institutionalization, as well as the type of institution entered, e.g., nursing home, hospital, jail, and military.

NOTES

1. The problem is diminished but not eliminated, since the first interview does not establish when a spell of unemployment began.

2. It might be more accurate to obtain pension benefits from the pension formulas used, but pension formulas are too complicated for it to be feasible for researchers to use them to work out the pension benefits for a national sample of workers drawn from thousands of firms.

References

Burkhauser, Richard V. (1979) The pension acceptance decision of older workers. Journal of

    Human Resources 14:63-75.

Burtless, Gary, and Hausman, Jerry (1982) Double dipping: The combined effects of social

    security and civil service pensions on employee retirement. Journal of Public Economics

    18:139-159.

Chamberlain, Gary (1982) Panel data. SSRI Working Paper #8209, University of Wisconsin.

Dahmann, Donald C. (1984) Panel surveys as a source of migration data. Proceedings of the

    American Statistical Association, and comments by M. David, pp. 517-522; 534—536.

David, Martin (1985) Distribution of income in the US: Implications for the design of SIPP. This

    issue, Journal of Economic and Social Measurement.

Fields, Gary S. (1976) Labor force migration, unemployment and job turnover. Review of

    Economics and Statistics LVII1:407-415.

Fields, Gary S., and Mitchell, Olivia S. (1984) Retirement, Pensions, and Social Security.

    Cambridge, Massachusetts: MIT Press.

Flinn, Christopher, and Heckman, James (1982) New methods for analyzing structural models of

    labor force dynamics. Journal of Econometrics 18:115-168.

Haber, Sheldon E. (1985) A Perspective on Linking SIPP to Administrative and Statistical.

    Records. This issue, Journal of Economic and Social Measurement.

Hausman, Jerry (1981) Labor supply. How Taxes Affect Economic Behavior. H. Aaron and J.

    Pechman, eds., Washington, D.C.: Brookings Institution.

Heckman, James, and Singer, Burton (1984) A method for minimizing the impact of

    distributional assumptions in econometric models for duration data. Econometrica

    52:271-320.

Jakubson, George (1983) Unobserved Heterogeneity in Labor Economics. Unpublished Ph.D.

    dissertation, Department of Economics, University of Wisconsin.

Joreskog, K.G., and Goldberger, A.S. (1975) Estimation of a model with multiple indicators and

    multiple causes for a single latent variable. Journal of the American Statistical

    Association 10:631-639.

Kaitz, H.B. (1970) Analyzing the lengths of spells of unemployment. Monthly Labor Review

    93:11-20.

Katz, A., ed. (1977) The economics of unemployment insurance: A symposium. Industrial and

    Labor Relations Review 30 (July).

Kiefer, Nicholas, and Neumann, George (1979) An empirical job search model with a test of the

    constant reservation wage hypothesis. Journal of Political Economy 87:89-108.

Moffitt, Robert, and Nicholson, Walter (1982) The effect of unemployment insurance on

    unemployment: The case of Federal supplemental benefits. Review of Economics and

    Statistics LXIV.1—11.