LABOUR AND INCOME Dynamics

Volume 6 Number 2 May 1997 (Product number 75-002-XPB)

WAVE 2 PUBLIC-USE MICRODATA FILES

The Wave 1 release contained two public-use microdata files: a PERSON file and a JOB file. The PERSON file was a longitudinal file as well as a cross-sectional file in that it contained some 1992 data (from the January 1993 paper preliminary interview) in addition to 1993 data. So some records had a zero cross-sectional weight and others had a zero longitudinal weight. The JOB file was essentially the same as the PERSON file; the difference being one record per job (with non-job level data repeated for every job held by the person) as opposed to one record per person (with job data strung out on the same person record). No linkage of persons in the same household or same family was allowed so a section of data on Other household members (OHM) was added to every record on PERSON. Up to 10 OHM's were allowed.

The basic intention is to replicate the wave 1 approach for wave 2. However, certain issues have arisen, mostly as a result of comments received from data users, which have led to certain changes being considered for wave 2.

Editor's Note

Many of you will be attending this year's Learned Societies meetings in St. John's. SLID has accepted an invitation to give a half-day workshop for their data users. This workshop will be held on June 9 in the afternoon. SLID data users should find that this date coincides with their society's meetings. More information is given in a separate note in this issue of Dynamics. Although you may have already started using the data files, I hope that you will have time to attend this workshop.

Within the next few months, the SLID wave 2 public-use microdata files will be available. The cost for this CD-ROM will be $1000 for those who purchased the wave 1 product, and $2700 for those who are purchasing SLID microdata files for the first time. You can save time later by paying now and signing the Statistics Canada microdata license agreement. In this way, you will be sent the CD-ROM as soon as it is available.

Nathalie Noreau

Statistics Canada Statistique Canada

Canada

Differentiation between cross-sectional and longitudinal files

With two years of data, the cross-sectional use of a combined file would be much more complex than if separate cross-sectional and longitudinal files were created. Therefore, starting with wave 2, separate files will be released annually. To compensate for the different approach used for the wave 1 release, the wave 2 release will contain 1993 and 1994 cross-sectional files. It must be stressed that it is not possible to produce a cross-sectional file from the longitudinal file as there are persons with zero longitudinal weight (and therefore would not be on the longitudinal file) who have a non-zero cross-sectional weight.

OHM data

The question of whom to include as OHMs was straightforward with one year of data, all persons in the same household on December 31, 1993. With two years of data, this would be extended to all persons in the same household on either December 31, 1993 or December 31, 1994. In addition, 1993 and 1994 data for the OHM would be required. On a flat-file, this would require a significant amount of space, which would increase each year. There was also the question of how usable this type of format would be. So OHM data will be provided only through a special request and will not be part of the regular wave 2 release.

Relationship between PERSON file and JOB file

For wave 1, the PERSON file and JOB file were designed to be independent of each other. Essentially, all variables were included on each of the files. The number of variables and record lengths of these files were: 996 and 3205 for the PERSON file and, 496 and 1482 for the JOB file. Multiplying these numbers by 6 to give a rough estimate of the sizes of the six-year files indicate that this approach is probably not feasible in the long-term.

Even in the short-term, the files are quite complex, and a different approach is planned. Basically, person-level variables would be found only on the PERSON file and job-level variables would be found only on the JOB file. To use variables which are on separate

files would require linkage between the two files using the person identifier. To eliminate some situations where linkage would be necessary, the JOB file would contain some person-level data (for example, age, sex, highest level of education, years of work experience, marital status, family size, family type, family income).

Content of cross-sectional files

The cross-sectional files would have a structure analogous to that of the longitudinal files (i.e., person-level variables on PERSON file and job-level variables on JOB file, with some overlap of variables to reduce the number of file linkages). In terms of variables, the basic approach is to include only variables which pertain to the reference year of the file. So the following variables are included:

- those which are only collected once (such as date of birth, sex)
- those which are collected annually (such as marital status, years of schooling)
- spell data covering only the reference year (such as jobs, jobless spells, and job absences)

However, marital spells are not included on cross-sectional files.

Suppressed variables

For wave 1, variables which were available internally, but which were suppressed on the public-use files for confidentiality reasons, were included on the public-use file record layout and "Not applicable" values included on the file. The purpose of this was to indicate to users the complete range of variables available from the survey. It also allowed users to submit programs to access the internal database as the record layout was identical for the two files. To reduce the size of the public-use files, it is proposed to not include these variables on future releases. However, the complete variable descriptions will still be included in the data dictionary, so users will still have descriptions of all variables.

DATA USES: LABOUR MARKET PERFORMANCE OF FOREIGN- AND CANADIAN-BORN

Conducted by Derrick Thomas, a Statistics Canada analyst in the Housing, Family and Social Statistics Division (613-951-2093 or thomder@statcan.ca), this study will apply socio demographic analysis techniques to identify working and earning differentials of the foreign born in relation to those of the Canadian born. The SLID sample of almost 30,000 participants also includes over 3000 respondents with an immigration date to Canada.

Selected cross-sectional and dynamic dimensions of job market attachment and mobility will be analysed in terms of the following labour market indicators:

- Unemployment and job interruption contrasts
- Comparison of earnings differentials
- Patterns of job separations and re-entry into the labour force

The analysis will provide insight into the dynamics of labour market, where differentials, such as job duration and job mobility occur. It will assess the importance of socio-demographic characteristics such as age, gender, and immigration status.

The foreign born will also be profiled by length of stay in Canada, official language ability, level of education, country of education, visible minority status, and place of birth. These may be important covariates of time dependant employment and occupation differentials between Canadian and foreign born population.

An examination of perceived barriers to employment. The conclusion will identify implications for training programs that enhance employment opportunities for all Canadians.

DATA USES: TRANSMITTING EDUCATION AND SOCIO-ECONOMIC STATUS FROM PARENTS TO CHILDREN

Fernando Mata, Department of Canadian Heritage (613-997-7910 or fernando_mata@pch.gc.ca), is using SLID data to examine the hypothesis that a person's education and labour force outcomes may be originally traced to influences related to the parents' education. Schooling is seen as a critical mediating variable between the parents' educational levels and the child's eventual participation in the labour force, occupational status and earnings levels. Labour force members of different visible minority status, birthplaces, mother tongues and ethnicity-related characteristics are being studied. A path analytic model of parental educational transmission is being fitted. Initial testing of this model across different ethnicity segments strongly suggests that parental education, by affecting directly the child's schooling, also influences the child's labour force participation, occupational status and earnings levels. However, both the specifics of the parental educational transmission and the main agent of transmission (father or mother) varies substantially from group to group. The results of this research will be documented in an upcoming ILDS working paper.

THE GEOGRAPHY OF SLID

As for many other aspects of longitudinal surveys, geographical variables are more complicated than they are with "traditional" cross-sectional surveys. In the Labour Force Survey (LFS), for example, all locational information for respondents is drawn from information on the sampling frame. Since dwellings and not persons are sampled for the LFS, movers are replaced in the sample by the new occupants of the dwelling. In SLID, respondents are followed when they move, and information on their new residence is collected. In general, SLID provides geographical information for their place of residence on December 31 each year. Users can analyse movers longitudinally, including them in their new geographic areas for the purposes of geography-specific tabulations. SLID also generates mobility variables which indicate whether a respondent lives in a different province, metropolitan area, or municipality compared to the previous reference year.

Linkage to the many levels of geographic areas is made possible through the postal code. Each respondent reports the postal code of his or her current residence. Processing of the survey data includes the use of Statistics Canada's Postal Code Conversion File (PCCF), which links every postal code in Canada to higher-level geographic codes. These codes are then added to the SLID database so that each record contains information such as the municipality, the county, and the federal electoral district as well as the Employment Insurance Region, and latitude-longitude coordinates for distance calculations. The postal code is a good choice as a "building block" since virtually all respondents know theirs, and because it is almost always small enough to determine unambiguously all other geographic areas of interest. Unlike the Wave 1 file, processing steps have been implemented to ensure that everyone now has a valid postal code and related geographical variables.

Although it may not be immediately obvious, the various SLID variables related to the LICOs (low-income cutoffs) depend on the assigned geography. LICOs are defined according to the family size and the "size of area of residence." The Survey of Consumer Finances (SCF) derives the LICOs annually using a definition of "size of area of residence" which differs from that used on the PCCF. The basic difference is that the size is determined for the SCF using a much broader area. For example, all persons within the same census metropolitan area would be assigned the same size for the SCF but not necessarily for the PCCF. Since the use of information on low-income is an important aspect of SLID, a second "size of area of residence" is derived to make comparisons with the SCF more meaningful.

More details on SLID geography and its impact on low income measurement will be available in a forthcoming SLID working paper.

SLID WORKSHOP AT 1997 CONGRESS OF LEARNED SOCIETIES

As part of the 1997 Congress of Learned Societies in St. John's (Newfoundland), a workshop will be presented for registrants.

Date: June 9, 1997
Time: 1:00 to 4:00
Place: Room 1009, Computer Sciences bldg
Memorial University of Newfoundland

Given by survey staff, this workshop is intended for persons who require a working knowledge of the survey. It will be of interest to the following wide range of persons:
Those with little or no knowledge of the survey who would like to determine its usefulness and relevance to their work;
Those who have determined that the survey data would be useful to them, but require more information in order to proceed with using the data;
Those retrieving SLID data for others, such as research assistants.

The duration of the workshop is one-half day, and covers the following topics:
Objectives of the survey
Highlights of the survey design
Data content
Relevant methodological issues
Potential areas of research
Survey data products

There is no charge to attend, but register early as space is limited. Use Section 3 on the regular registration form (society number 99). To request a registration booklet, contact:
1997 Learned Societies Secretariat
2nd floor, Spencer Hall
Memorial University of Newfoundland
St. John's, Newfoundland A1C 5S7
Telephone: (709) 737-4360
Fax: (709) 737-4449 Internet: learneds@morgan.ucs.mun.ca

WORKING PAPER SUMMARIES

The following are recently released working papers which can be ordered individually ($5) or by annual subscription ($25 on diskette or $50 for paper versions for 12 to 15 papers).

97-01 SLID questionnaire for demographics and contact: 1997
Ruth Dibbs, Debbie Lutz, Robert Kaminsky

This is the standard "print" version of one of the four annual SLID questionnaires. The "demographics" and "contact" modules are attached to the beginning of both the January and May interviews.

97-02 Differences in income estimates for persons
Harry Champion

97-03 Differences in income estimates for families
Harry Champion

These two working papers compare the income estimates produced by various data sources. Where differences occur, an examination of their causes is included. These papers provide a useful consolidation of background information on sources of income data.

97-04 Selection of a top-up sample for cross-sectional income estimates
Michel Latouche, Sylvie Michaud, Johanne Tremblay, Ruth Dibbs

Once data collection for SLID and the Survey of Consumer Finances (SCF) is integrated, an annual sample will be required to enable us to produce income data which has the same reliability as that currently being produced by SCF. This working paper outlines the steps taken to determine how this annual top-up sample should be selected.

97-05 1997 Preliminary interview questionnaire
Ruth Dibbs, Debbie Lutz, Bob Kaminsky

This is the "print" version of another SLID questionnaire. The preliminary interview is conducted as part of the January interview for those who are new to the SLID sample at that time and for existing sample members who turned 15 during the reference year. The 1997 preliminary interview was conducted on a much larger sample as compared to the few previous years as the January 1997 interview saw the first interviews for our Panel 2 respondents.

97-06 Survey of Labour and income Dynamics: Processing Strategy for Wave 1 Income Data
Élaine Fournier, Tracey Leesti, Mylène Lavigne

From the point of view of the SLID team, there is a distinct difference between labour and income data. The labour content includes many variables with very little processing required for most variables as the impact of small errors is not significant. On the other hand, there are relatively few income variables. It is a very sensitive topic for survey collection. Even for those who are willing to report income information, it is difficult for many respondents to do so. SLID also collects income data in two very different ways: through a traditional interview and by accessing administrative files. Some processing of income data is required as the impact of small errors can be quite significant. This working paper outlines the processing steps followed for the processing of the first year of SLID income data.