LABOUR AND INCOME

Dynamics

Volume 4 Number 3 **September 1995** (Product number 75-002)

$$$$$$$$$$$$$$$$$$$$$$

STRATEGY FOR PROCESSING

One of the current “hotbeds” of activity for SLID staff is processing of Wave 1 data collected in 1994. This has taken us longer than expected, as we have taken a long term view. However, the foundation has been laid for all future processing. We feel that this will provide us more timely data releases down the road.

The centerpiece of this approach is the development of the data model) a structure for organizing and storing the survey data. The SLID data model will handle survey data covering multiple years of data for multiple panels of respondents in one database. This design stage need not be repeated until major content changes are introduced.

The master database, accessible only by those within Statistics Canada, will be in a relational database structure. In addition to data intended for output and analysis, the database will contain information on the sample required for future data collection, performance indicators from past collection periods, and information required for the calculation of sampling weights and variances. Public-use microdata files will be a subset of the internal master database.

Created from the collected data, the SLID master database will reside on a dedicated server. Processing will take place in this PC environment. SLID processing can be described as being composed of the following steps:

Editor's Note

Starting with this issue, I’m pleased to take over from Philip Giles as editor of *Dynamics*. Meanwhile, Philip is never far, and in fact he is still your first contact for issues related to analysis and dissemination of SLID data. To prove the point, my Internet address is: giles@statcan.ca. The mailing address at SLID is still 11-D8, Jean Talon Building, Tunney’s Pasture, Ottawa, Ontario, K1A 0T6, and the fax number is (613)951-3253. My phone number is (613)951-4353.

I’ll be happy to hear from you any time about suggestions or questions regarding SLID news, including topics you would like to read about in our newsletter.

Heather Lathe

Statistics Canada Statistique Canada

Canada

preprocessing, loading, coding, editing and imputation.

- *Preprocessing:* This is a set of primarily manual steps. It involves the removal of duplicate data for the same person resulting from interviewer errors and transmission difficulties; ensuring that changes in household composition and move dates are consistent; reviewing interviewer notes which may lead to data corrections.

- *Loading:* This is the process of moving and converting the collected question responses into the variables in the internal master database. Due to complex question flows and the many derived variables, this represents a large component of the initial processing development.

- *Coding:* Several fields collect textual information which must be translated into a numeric code before being useful for analysis. The assignment of these codes is a mixture of automated and manual procedures. The codes are then "loaded" into the database.

- *Editing:* Compared with traditional data collection methods, the use of computer-assisted interviewing allows much of the editing to be done during collection. Flow problems, where the interviewer does not follow question skips properly, will be almost entirely eliminated. Some major consistency edits have been incorporated in the collection software, but not all inconsistencies are resolved during collection. Editing is one area where an incremental approach is being applied rigorously. Only those edits deemed essential will be included initially. Over time, as data problems emerge and greater knowledge is gained about the data, additional edits will be developed.
  One key element in the editing strategy is that all editing is preliminary until data collection for a panel is finished. That is, changes may result after further data are collected.

- *Imputation:* This is the process of assigning a value to variables which do not have a valid value, due to nonresponse at collection or to valid values suppressed as a result of data inconsistencies. For the first data files, only a certain number of key fields will have missing values imputed.

To the extent possible, manual processing is avoided. This greatly reduces the time required for processing. As well, it is much easier to document the nature and impacts of processing decisions. For automated processing, programmers and subject matter staff are collaborating, following predefined steps. A key aspect of this approach is the systematic and thorough documentation of the detailed processing specifications in electronic form. A more detailed description of the development and implementation of processing specifications will be provided in a future SLID research paper.

$$$$$$$$$$$$$$$$$$$$$$

PUBLIC-USE MICRODATA FILES

Although no specific date has been determined, SLID is committed to releasing its first public-use microdata file by the end of 1995. More information will be provided in the next issue of *Dynamics*. To help users plan for the release, an overview is provided in this issue.

The first file will include data collected from SLID's first panel of respondents in the January 1993 preliminary interview and in the 1994 annual labour and income interviews. In general, the data refer to reference year 1993. The data file will be stored in a flat rectangular ASCII format.

To help users prepare, we are producing a SLID research paper describing in draft form the content of the public-use microdata file. To inform users about variables on the internal database which are not being included on the public-use file, documentation of the internal database will also be available. Subscribers to the SLID research paper series will receive these documents automatically. Others may order copies if they wish to receive them prior to purchasing the microdata file (see "Research Papers" below).

$$$$$$$$$$$$$$$$$$$$$$

THINK LONGITUDINALLY!

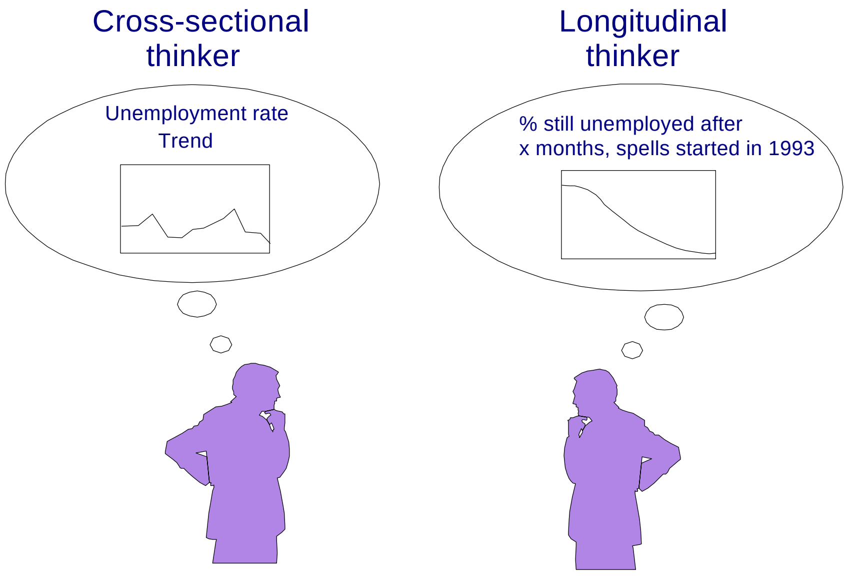
With several new surveys set to produce panel data from Statistics Canada over the next few years, analysts here and elsewhere have a definite challenge to start thinking longitudinally as well as cross-sectionally. The new data sources will give a more accurate picture of events that can only be inferred from trend analysis of cross-sectional data.

To give an example of the difference between trend analysis and longitudinal analysis, suppose that we were to observe a decline in the number of people holding unionized jobs (a cross-sectional analysis). From there, we may have some idea that job separations from unionized jobs have increased or, alternatively, that fewer people are being hired for unionized jobs. Adding the longitudinal analysis would show the relative importance of each.

Study the "event"

The switch from trend to longitudinal thinking is simple if we make the individual's *event* the focus of analysis. For example, rather than track people who are unemployed, the study population could become "people who lose their job" over a specified time. Longitudinal data give a "before and after" picture from which to study the determinants or the consequences of job loss.

Other events that could be examined using SLID data include changing jobs, returning to school full-time, getting off social assistance, or forming a common-law union, to give just a few examples. The variability of “states” that people experience over time may also be of interest. For example, what proportion of the population experiences relatively stable income over time?



Solve the “moving” target

This is a question of defining the target population for a single point in time, for example persons who married or had a child *in 1993*. Unlike cross-sectional analysis, where the people making up the population change according to who fits the criterion each year, longitudinal analysis looks at the same people over time, based on whether they fit the criterion at the selected time. The population can be fixed at either the beginning or the end of the total period studied, depending on whether a “before” or “after” picture is required.

The definition of the target population indicates which year’s longitudinal weight should be used, since the theoretically constant longitudinal weight is in fact adjusted slightly for non-response each year.

Use the individual when studying families

For longitudinal analysis, the same logic requiring that you study the same individuals over time would suggest that the same families have to be tracked over time for family data. However, families can change composition over time and do very frequently, causing great difficulty for the definition of “longitudinal families”. The solution is very simple: study the family characteristics of individuals. The individual becomes the unit of analysis, and family characteristics are attributes of the individual.

$$$$$$$$$$$$$$$$$$$$$$$

REGIONAL ANALYSIS: A SPECIAL CASE OF THE MOVING TARGET

Not long ago, a prospective data user asked, “Is it possible to get the microdata file for just one province?” Here is an explanation of why a longitudinal analysis of one province requires the full dataset.

For example, at the start of the period of analysis, we define the population as those people living in British Columbia and observe their characteristics. At a later date or reference period, we observe their characteristics again -- by this time, they may be dispersed throughout Canada and abroad. Alternatively, we could define the population as at the *end* of the period of analysis and observe the *earlier* characteristics of those people, wherever they were located.

The implication is that even for analysis of a specific region, the dataset must be allowed to encompass all of Canada (and overseas). This is also the rationale behind our intention to collect information from longitudinal respondents who move to the territories or the United States.

In a few cases, the population logically could be defined as those people living in British Columbia at one time and still living in the province at a later time. However, when a subset of the population is excluded this way (specifically, those who move outside or into the province within a given time period), there is always the possibility of a bias in the characteristics compared to what would be obtained with the larger population.

Geographical detail

To maintain confidentiality of the public-use file, limited geographical information will be available on the file. Finer levels of geographic detail will be available through custom requests, subject to sample size constraints and confidentiality of the final results. Geographic detail in SLID includes Economic Regions, Census Metropolitan Areas, and any other area which can be defined using postal codes. It also includes the approximate size of area of residence (part of the urban/rural variable). The second panel of SLID, which will double the sample size, will provide greater flexibility in defining geographical areas.

$$$$$$$$$$$$$$$$$$$$$$$

RESULTS OF THE INCOME TAX PERMISSION QUESTION

In the last issue of *Dynamics*, we described the process for asking respondents their permission to access their tax records for income information. The advantage for respondents is that they do not need to complete the income interview. The advantage for the survey is increased precision in income reporting and potentially less loss through attrition.

A large proportion of respondents chose the tax route in May 1995 -- 63%. SLID will use tax file information for them if the match to their records is successful. Some people said they did not file a return, and 31% said no to the tax permission question.

$$$$$$$$$$$$$$$$$$$$$$$

RESEARCH PAPERS

The following are recently released Research Papers which can be ordered individually ($5) or by annual subscription ($15 on diskette or $50 for paper versions). For more information, contact Anne Palmer by phone at (613) 951-2903, by fax at (613) 951-3253, or by mail at 11-D8 Jean Talon Building, Tunney's Pasture, Ottawa, K1A 0T6. Internet users: giles@statcan.ca.

95-06 Dependent interviewing: Impact on recall and on labour market transitions
Alison Hale and Sylvie Michaud

Feeding back information reported at the previous interview is done for certain characteristics to aid respondents' recall, leading to increased data quality. This report examines the extent to which dependent interviewing was effective for the January 1994 SLID labour interview.

95-07 Some Effects of Computer-assisted Interviewing on the Data Quality of The Survey of Labour and Income Dynamics
Ruth Dibbs, Alison Hale, Robert Loverock, Sylvie Michaud

With computer-assisted interviewing (CAI), one can conduct more complex consistency checks in the field than is possible with the traditional paper-and-pencil interview. The result is reduced response errors. This report examines expectations regarding the quality of data in three content areas: labour force activity, respondent-sensitive sources of income, and relationships between household members.

95-08 The Survey of Labour and Income Dynamics: Visible Minorities and Aboriginal Peoples
Ruth Dibbs and Tracey Leesti

Funded by the Interdepartmental Working Group on Employment Equity Data, this report evaluates the extent to which SLID data may be used for employment equity purposes. Particular focus is placed on two of the four employment equity designated groups: visible minorities and aboriginal peoples. Comparisons to census counts are included.

95-09 SLID Derived Variables: Educational
Heather Lathe, Philip Giles, Joanne Murray

Frequently, the information needed by data users cannot be reliably obtained by a single question collected from respondents. Thus, data files produced by surveys contain many "derived variables", those not collected directly from respondents, but calculated or derived from survey responses. SLID intends to document all derived variables, including definitions of how they are derived. This research paper is the second such document, focusing on educational variables.

95-10 Graphical description of SLID content
Philip Giles

SLID collects a wide range of information. To aid users of the data, this content has been described in many ways. This document provides the content description in a different format, namely graphically. It is detailed enough to give users a feel for the range of information, but does not provide detail on data variables.

95-11 Disseminating Data From Longitudinal Surveys: Issues Facing the Survey of Labour and Income Dynamics
Maryanne Webber

The dissemination of microdata from longitudinal surveys poses several challenges. The purpose of this paper is to outline these challenges and some of the measures being proposed to deal with them. The main purpose of the paper is to provoke discussion on general dissemination issues, using SLID as a case study.

95-12 Questionnaire and collection procedures for SLID income data collection - May 1995
Élaine Fournier, Alison Hale and Bob Kaminsky

In May 1995, for the second time, SLID collected annual data on income from its first panel of respondents. This paper describes the collection procedures and provides an overview of the interview process.