



**INTEGRATION OF SLID AND
THE SURVEY OF CONSUMER
FINANCES (SCF)**

Conducted each April as a supplement to the Labour Force Survey (LFS), the Survey of Consumer Finances (SCF) has traditionally been the source of Statistics Canada's annual income estimates, including statistics on low-income. The survey collects income information by source pertaining to the previous calendar year and uses labour and demographic information from the LFS to increase the analytical richness of the data.

In April 1997, the SCF will collect data for the last time. Instead, the SLID sample will be used for production of the annual, or cross-sectional, income estimates, starting with reference year 1997 data, in addition to continuing the production of longitudinal income data.

There are two reasons for integrating. First, there is sufficient overlap in the survey objectives and content of the two programs to allow for cost reduction, or more effective use of resources,

Editor's Note

Not only does this issue begin another year of Dynamics, but it brings a change in scope. Articles in this issue outline the integration of SLID with the Survey of Consumer Finances, Statistics Canada's annual survey for producing income estimates including low income rates. As a result, Dynamics will now be bringing you

information on the integrated program. Now known as the Income and Labour Dynamics Working Paper Series, the research paper series will have a similar expansion of scope.

Nathalie Noreau


through a merging. Second, integration will promote data harmonization. Without it, for example, income estimates will differ for reasons that are difficult to quantify and to explain to data users.

Content issues

The integration of the demographic, labour and income content from SLID, and the equivalent in the SCF, appears quite feasible. The contextual information available for the analysis of cross-sectional income data will be enhanced by this move, because the demographic, cultural and labour market information will be drawn from the current SLID labour interview rather than the LFS. Also, as the former uses an annual reference period, it is inherently more suitable as a companion for annual income data. Some priority areas where the data will be enriched are the earnings of men and women, which can provide more detail on those who are not full-year, full-time workers; family dynamics data, which will allow us to link family changes to family income data; work experience and expanded education data, to use in analysing wages and salaries; spells of receipt of such government transfers as unemployment insurance, worker's compensation and social assistance (rather than just an annual dollar amount).

Cross-sectional time series consistency

The shift can be expected to cause a break in the cross-sectional time series. Perhaps the most important cause is that SLID offers respondents the option of providing access to their tax file data, in order to reduce attrition and response error. So far, in the first panel, over half of the respondents are providing income data via the "tax route" rather than the "interview route". This option will become a feature of the cross-sectional data as well, starting with this year's collection. After the data are collected, the two collection routes are merged into a single dataset that comprises a series of reasonably equivalent income categories. From the studies



done to date, it is clear (and not surprising) that the reporting patterns in the two sources are not identical. Thus, the use of “tax route” data will probably cause a time series rupture. However, comparisons of SCF and SLID data over the next two years will provide some advance notice of the expected impact of this important factor.

Timeliness

The use of tax file data also has an impact on timeliness. Currently, the first SCF income estimates are released in November, followed by a series of releases over a period of several months. With the use of tax data, the first release of cross-sectional income data cannot be that early, because the Revenue Canada Taxation file is itself not available until the fall. On the other hand, the closer alliance with tax file data will probably mean that the release schedule for a given year is much more compressed. The goal is to deliver the data within 15 months of the end of the reference year.

Sample design and estimation

The key feature of integration is sample pooling: the same respondents will provide both cross-sectional and longitudinal data. The longitudinal sample will be augmented each year by a top-up sample to maintain cross-sectional representativity. Each of the two longitudinal panels is about 15,000 households, yielding a sample of about 30,000 households on a yearly basis. Once attrition is taken into account, a top-up of about 10,000 households per year is needed to maintain the current reliability level of cross-sectional income data. Those in the “top-up” sample will be asked the SLID preliminary interview and labour interview in January. Following standard procedures now in place for SLID, “top-up” sample members will then choose to allow the use of administrative income data or to provide income data in a May interview. The adherence to the same collection procedures used for the “regular”

sample members will allow the integration of data for the two sample groups into one database.

Data collection and respondent relations


Integration presents the combined program with an opportunity to review data collection procedures and the approach to respondent relations. With a few years of experience with computer-assisted interviewing, this is an opportunity to step back and consider how the collection application can be improved -- for example, to reduce reporting and keying errors. Income is a sensitive topic in the field, but we can perhaps make it a more “approachable” topic for respondents and interviewers. Because the merger will in any case affect cross-sectional time series continuity, it affords a chance for innovation in data collection techniques and respondent relation approaches.

Processing, edit and imputation

The SLID processing system will continue its development path, although program integration will have an effect, as cross-sectional data assume greater importance. The SLID system will incorporate SCF “best practices”. The presence of longitudinal data for a majority of respondents and the use of administrative income tax data will play a fundamental role in the data processing.

Products and services


The need for cross-sectional and longitudinal data products will continue, as well as products which show how the two types of data complement each other. In the short term, the current product line from the two programs will be continued. In the long term, consultation with data users on changes to the product line are planned, although certain directions in place for Statistics Canada as a whole will be followed. Among these are: a shift away from paper products; harmonization of data, product “look and feel” and user documentation; enhancement of the analytical content of



output program; and collaborative analysis undertakings with other parts of Statistics Canada as well as with those outside Statistics Canada. A major concern in the sphere of dissemination is the impact of sample integration on the confidentiality of microdata files. Currently, public-use microdata files produced from SCF allow linkage of persons within households. Part of the confidentiality protection for SLID public-use microdata files is to suppress such linkages.



THE INTEGRATION “BRIDGE”



The initial plan was to use the SLID production system for cross-sectional income data when the switch in collection was made (for reference year 1997 data). However the SLID system has not yet stabilized and is in a “catch-up” mode. At that time, it will not be in a position to produce cross-sectional income data that meet the user community’s expectations for timeliness.

Therefore, for the 1997 and 1998 income years, an interim processing system (called the “bridge”) will be developed to produce timely cross-sectional income data. The input to the “bridge” will be a combination of edited and unedited data from the longitudinal sample and the unedited data from the top-up sample. To the extent possible, the bridge will use a combination of existing code from the current SCF and SLID production systems. Only a limited number of income, demographic and labour variables will be selected for use in the “bridge”. The variables selected are those required to produce the traditional “high-profile” income statistics, including average incomes, low income rates, female/male earnings ratios, and the effects of transfers and taxes on family incomes. The 1997 income data processed by the bridge system is planned for release in March 1999. This fifteen month difference between the

end of the reference year and data release will continue with the conversion to an integrated single program production system.


DATA ON HOUSING AND HOUSEHOLD EQUIPMENT

The current Income and Housing Statistics program covered the SCF as well as the annual Household Facilities and Equipment Survey (HFE); the annual Repair and Renovation Survey, now a full cost-recovery survey; and occasional ad hoc surveys. Although done a month later, there is sample overlap with the SCF, which allows the use of total and detailed income data in analysing housing and other characteristics. Indeed, the HFE, SCF and rent data from the LFS are combined into an annual database that has proved to be very popular.

The housing and household equipment information will continue to be collected, but as part of the Family Expenditures Survey (FAMEX). The disadvantage of this change is the loss of detailed income data to use in conjunction with the HFE. However, it was felt that some of the HFE content is not appropriate for a longitudinal vehicle and thus adds response burden, arguably without sufficient benefits to warrant the increased risk. HFE also includes questions that are more oriented to “consumer goods” -- kitchen and household appliances, for example. For these items, the value of a longitudinal perspective is unclear, and there is clearly a better “fit” with FAMEX.

EQUIVALENCE SCALES

For the purposes of identifying low income families, it is generally agreed that the size and composition of each family should be considered. For example, a single person at a given income level should be less likely to be counted as having a low income than a family of five with the same dollar income.




Equivalence scales are used as an index of the needs of a given family, relative to a single person family. In this way, family incomes can be compared across all families by examining an “adjusted” family income: the actual family income divided by the equivalence scale.

There is no generally accepted equivalence scale, as ultimately some subjective judgement is necessary. The wave 1 SLID file included an equivalence scale used by the OECD (Organization for Economic and Cooperative Development). This is not the equivalence scale used by Statistics Canada for the production of its LIMs (low income measures). This second equivalence scale will be included on future SLID public-use microdata files.



NOTES OF ADVICE FOR
USERS OF THE PUBLIC-USE
MICRODATA FILE



Of the many comments and questions we have received since the release of the wave 1 SLID public-use microdata file and CD-ROM product, some commonalities have occurred. This article is intended to stress a few important points.

Reserved codes

Although they are not always used, all SLID variables have four “reserved codes”, which have a special meaning. Users must take consideration of these reserved codes, particularly for numeric variables. With a few exceptions, the reserved codes are the highest four values for field, dependent on the length of the field. For a field of length 1, these values are 6, 7, 8, 9. For a field of length 2, these values are 96, 97, 98, 99, and so on for fields of greater length. The meaning of these codes is:
6, 96, 996, ... = Not in SLID sample (This applies to sample members who are not there for all survey years.)

7, 97, 997, ... = Don't know (In general, this means that we do not know the value, but some of these are a result of the respondent not knowing.)

8, 98, 998, ... = Refusal (Person refuses to provide the value.)

9, 99, 999, ... = Not applicable (Person was not eligible for the data item.)

Do not perform arithmetic operations, such as calculating means, without removing reserved codes.

Record constraints in IVISION

Following on the previous discussion, some users have expressed frustration over the limited functionality of the record constraint when using the IVISION Browser software. Future versions of the software will allow greater flexibility. In the meantime, a way to retrieve a table with a complex record constraint is to derive a new variable (using one of define recode, define bands, or define derived fields). Use this new variable as part of the table definition, then focus on that section for the table related to the population of interest.

Screening for population aged 16 to 69

The microdata file includes a record for all persons aged 16 or over at the end of 1993, the reference period for most of the data. The labour data is only collected for persons aged 16 to 69, due to the very low labour market participation rate of those aged 70 or more. Unfortunately, it is not obvious how one screens for the population aged 16 to 69 on the public-use microdata file. This oversight will be corrected on future files. For the wave 1 file, the "trick" to screening for the population aged 16 to 69 is to use only records with values of variable STUDEF26b different from 9. [This variable is an indicator of whether the person was a student in the reference year. However, this is irrelevant to the "trick".]




RESEARCH PAPER
SUMMARIES

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

96-09 Educational Attainment: A Key to Autonomy and Authority in the Workplace

George Butlin and Jillian Oderkirk

Results from analysis of data from the 1993 Survey of Labour and Income Dynamics and the 1994 General Social Survey, indicate that, with few exceptions, education is one of the strongest predictors of an individual's ability to access occupations offering autonomy and authority in the workplace. This remains true, even after the effects of factors which also influence access to these types of positions, such as gender, age, firm size, years of work experience and industry, are taken into consideration.

96-11 Family Data From the Survey of Labour and Income Dynamics: 1996 Status

Philip Giles

A person's "family situation" is often an important consideration to various decisions. For example, a person may choose not to work in the paid labour market as his/her spouse may earn a sufficient amount for their requirements. Thus, family variables are important to many analyses, and this is even more true for longitudinal analyses. The use of longitudinal family data is complicated by the fact that a family can change over time as people move in and out, are born and die. The use of SLID family data is discussed, along with several examples to indicate how family data can be analysed longitudinally.