MEASURING INCOME DYNAMICS: The Experience of Canada's Survey of Labour and Income Dynamics

by

Maryanne Webber Statistics Canada Canada

for presentation at

Seminar on Poverty Statistics Santiago 7-9 May 1997

MEASURING INCOME DYNAMICS: The Experience of Canada's Survey of Labour and Income Dynamics

I. INTRODUCTION

In 1993, Statistics Canada launched a longitudinal household survey called the Survey of Labour and Income Dynamics (SLID). This survey is designed to track the labour market activities, family changes and income levels of Canadians over a period of time.

This was not Statistics Canada's first foray into the field of panel household surveys. Indeed, SLID had a predecessor, the Labour Market Activity Survey (LMAS), which was active from 1986 to 1991. LMAS was sponsored by Human Resources Development Canada, the federal government department responsible for labour market policy and for the administration of many labour market and social programs. LMAS data are still being "mined" by researchers. But there was a major content shortcoming: the lack of family income data.

Increasingly, analysts are concerned with the interrelationships between the labour market and the family's economic well-being. To some extent, we have witnessed a relaxing of the traditional barriers that existed between labour market research and research focused on family welfare issues. SLID's purpose is to provide a more holistic view, shedding light on the links between family, work and income. To do this, the survey needed to be longitudinal, and to track both family changes and income changes over time.

Statistics Canada has a long-established time series of low income data based on the annual Survey of Consumer Finances (SCF). These data are intensely used, and form the basis of several regular or ad hoc publications by groups concerned with poverty and income inequality. The existing cross-sectional data have been very useful for monitoring trends in income distribution, for understanding the variations by type of family and for analyzing the role of government transfers in maintaining income levels. However, the development of effective programs and measures requires a better understanding of the persistence of income inadequacy and of the events that trigger flows into and out of low income. At what stage should government programs intervene? What types of measures are the most fruitful?

The funding for SLID was obtained because several policy departments of the federal government recognized that this is an important data gap in the national statistical system, worthy of a stable long-term survey program.

This paper traces the history of SLID, from its development to today. It discusses some of the current challenges and lessons learned along the way.

II. SLID OVERVIEW

Because of its mandate to look at labour market behaviour, family circumstances and income, SLID carries a broad selection of variables. The labour market variables encompass information on the characteristics of jobs held during the year; for example, start and end dates, industry, occupation, hours worked, wage rate, union status, pension coverage, how the job was obtained and the reason for job loss. The dates of work absences are captured, along with the reasons for absence. During spells where no job was held, there are questions on job search and the desire for employment. This detailed information is summarized into higher-level variables that encapsulate the person's labour market patterns, including weekly labour force status (a value of employed, unemployed or not in the labour force for each week of the year), annual labour force status (a summary variable showing all the labour force states experienced during the year), the number of jobs held in the year, the total weeks of employment and unemployment, and so on. In addition, there is a range of family-level labour market variables summarizing the extent and nature of labour market involvement for the family as a whole.

The personal characteristics captured by the survey include basic demographics, educational activity, educational attainment, marital history, ethnocultural variables, such as immigration status and ethnic origin, and geographical mobility.

For income, the survey identifies about 20 specific sources of income, along with amounts. Also, total income before taxes, income taxes paid and after-tax income are recorded. Again, the amounts are available for both the individual and for the family as a whole. The income questions account for a small proportion of the total survey content, but the other variables are valuable in interpreting the data on income stability and adequacy.

Main survey design and data collection features

SLID is designed to be a continuous survey, but with a sample that changes over time. Each panel of about 15,000 households is selected for six years. An interview is conducted each year, in January, to collect and update information on labour market activities (over the previous year), educational activities and family changes. Then, in May, the household is contacted again to collect income information for the previous year -- unless the people in that household have agreed that we can access their tax records, assuming of course that they have completed a tax return. About 60%-70% of respondents have tax returns and agree to allow access. Otherwise, income information is collected via an interview. This is done in May rather than January to coincide with the time that most people have completed their tax returns and therefore have the relevant information close at hand. This is particularly important for people who are self-employed. Labour information is collected for persons

aged 16 to 69 and income is collected for persons 16 and over. Family information covers persons of all ages.

The individuals living in the dwellings selected at the beginning of a panel are called *longitudinal respondents*. These persons are followed for six years, whether they move or not, and regardless of age. In addition, the survey interviews *cohabitants*: people who share a dwelling with a longitudinal respondent at some time during the six years, although they were not initially selected for the sample. Cohabitants remain part of the sample as long as they continue to reside with a longitudinal respondent. This ensures that the family information collected on longitudinal respondents is up to date.

Unlike most Statistics Canada surveys, SLID is a voluntary survey. Although a six-year commitment on the part of respondents is significant, every effort has been made to maintain response burden at a tolerable level by keeping the interviews short. They take about one half-hour to complete per household.

The data processing challenge

Although difficult, the task of finalizing the questionnaire and collecting the first wave of data was not as demanding as the subsequent data processing stage. The survey's first reference year was 1993. It took roughly 24 months to process the first wave of data. The results for 1993 were released in June 1996.

The processing of the first wave was time-consuming because we wanted to build a database that would accommodate future "waves" of data. A longitudinal database must deal with the issue of how to represent time. Spells, events and transitions are the meat of a longitudinal survey. In general, the experience gained from building databases for cross-sectional surveys offers few clues on how to capture changes experienced over time.

Improving timeliness

While the results for 1993 were being processed, data collection continued each year. The 1994 results will be published in June 1996, which means that it will take 12 months to process the second wave, compared with 24 months for the first wave. When SLID is fully stable, data will be released within 15 months of the end of the reference year. However, because there is some "catching up" to do, we do not expect to reach this timeliness target until the 1999 reference year.

Dissemination focuses on microdata

So far, the main objective in data dissemination has been to make a public use microdata file available, and to ensure that our documentation is solid enough to allow researchers to do

their work. But this is just the tip of the iceberg. We have heard from many data users that the survey content is of great interest but they cannot (for reasons of time or technical training) work directly with the microdata. To ensure that there is a broad user base for the survey, it is therefore important to produce studies or tables that reflect transitions and flows -- that is, longitudinal rather than cross-sectional tables. The design of such tables is definitely a challenge, but it is one that we will be taking on in the coming year.

The protection of confidentiality is an important issue in the dissemination of public use microdata. With the first wave, the approach adopted was a combination of suppression of certain variables, rounding of income amounts, and collapsing of variables (such as age) into categories. However, as the amount of information on each person grows over time, the protection of confidentiality becomes increasingly difficult. There is an initiative underway at Statistics Canada to make microdata available through a "remote access" system. Essentially, researchers outside Statistics Canada would write and test their programs using a test file that we provide. The program would be submitted to Statistics Canada via Internet, executed against the full database, checked to ensure that the results are not problematic from a confidentiality perspective, and transmitted back to the researcher. Although simple in principle, the approach would require well-developed guidelines, fast turnaround and high caliber test files to work efficiently.

III. POVERTY DYNAMICS IN SLID: WHAT ARE WE LOOKING FOR IN THE DATA?

With the release of 1994 data in June, it will be possible to begin analyzing income dynamics using SLID. Two years is too short a period to look at the persistence of poverty but we can at least look at flows into and out of low income.

Currently, the most prominent measure in Canada is the Low Income Cut-off (LICO). As noted in the Appendix, LICOs vary by family size and size of community. There are 35 cut-offs in all. They are adjusted each year for changes in the cost of living.

LICOs refer to family income. The family in this case is defined as all persons living in the same dwelling and related by blood, marriage, common-law union or adoption. Family income is derived by summing the individual incomes of all persons belonging to the same family. Because a person's family composition can change over time, the approach we have adopted is to "freeze" family composition on December 31 of each year. Family income is thus the sum of the individual incomes of all people living together on December 31 of the year in question.

Since family composition can change over time, it would be difficult if not impossible to use the family as a unit of analysis in studies of income dynamics. The individual is a much more

manageable unit of analysis, but we are interested in a family characteristic, namely family income. Family income is therefore treated as a characteristic of all the individuals belonging to a given family. (This is not an unusual approach in cross-sectional low income studies in Canada. For example, estimates of child poverty are based on this notion of individuals classified by their family income.)

When the data for 1994 are available, we plan to look at flows into and out of low income using the following simple transition table:

	1994	
1993		
	Below LICO	Not below LICO
Below LICO		
Not below LICO		

The population to be included will be longitudinal respondents of all ages, classified by their family income level in the two years. This will give us an initial reading on the magnitude of these flows in relation to the population below the LICO in both years -- in itself new information.

Then, a number of possible reasons for change will be considered. One possible source is government transfer payments that may have been initiated or terminated. Another is change in market income, for example, someone in the family having started a new job or lost a job. A third important cause could be an event in the family itself -- for example, a family breakup, a youth returning to the home of his parents, and so on.

IV. DATA QUALITY ISSUES AND MEASURES TO ENHANCE QUALITY

Longitudinal household surveys share some of the quality problems of cross-sectional surveys, and they have additional concerns. In terms of quality, income is generally not a popular survey topic among respondents. The survey staff were concerned from the outset that, over time, refusals would increase to the point where the data would not be of acceptable quality. This concern was the main reason for offering respondents the option of providing their income information via tax records rather than an interview.

As with other income household surveys, there are concerns about unreported income sources, under-reporting of amounts, and under-representation of high-income households. To some extent, this can be evaluated by comparisons to external data sources. For example, there are aggregate estimates available for payments made through various government transfer programs and for total wages and salaries.

Every year, the SLID estimates will be evaluated against these external sources. The evaluation for 1993 resulted in some fine-tuning of the processing (for example, more rigorous editing of outlier values), and pinpointed areas for further study. Although the survey's goal is not to produce aggregate income estimates, these comparisons provide useful information on quality.

The quality concerns are not all on the income side of the survey. Recall is a potential problem in the collection of information on labour market activities, particularly among respondents who have complex activity patterns. Past retrospective surveys, looking at labour market activities over a full year, have shown that there is a tendency to forget short spells of unemployment occurring at the beginning of the year and also to "telescope" both employment and unemployment spells, that is, recall them as having occurred closer to the survey date than they really did. The effects of said errors include, for example, underestimation of unemployment and distortion of seasonal patterns.

Our main strategy for ensuring quality of the labour market information is to structure the interview in a way that assists the recall task and to feed back selected information from the previous interview. Regarding the interview structure, rather than proceed in chronological fashion, the first step is to identify the dates of jobs held. After the employment spells have been fixed, the interviewer then proceeds to "fill in the holes", identifying job search activities that occurred between employment spells. The rationale for this approach is that jobs are easier to recall and situate in time than job search activities.

Respondents are also fed back information on jobs in progress at the end of the previous year, which helps to reduce "seam problems". (Because of recall problems, surveys like SLID find a large number of employment and unemployment spells starting and ending at the "seam" between two reference periods.) Feeding back information has been shown to reduce this phenomenon.

Longitudinal surveys are particularly concerned about sample attrition -- a gradual erosion of the sample due to increased refusals and failure to trace respondents who have moved since the last interview. After three waves, SLID has retained about 81% of the respondents aged 16 and over who were contacted in Wave 1:

Panel 1:	Ν	%
Total longitudinal respondents 16+	30,900	100
Responded in all 3 waves	25,000	81
Did not respond in any wave	2,100	7
Dropped out after first wave	1,900	6
Dropped out after second wave	1,900	6

Non-response is caused by many factors. Outright refusals and failure to trace account for roughly half of all non-response. Results differ a little between January and May. In 1994, 25% of all non-response in January was due to refusals and 21% were unable to trace cases. In May, 28% were refusals and 16% were unable to trace. To put the results on failed tracing attempts into perspective, there were 2,700 respondents who moved in 1994, of whom 2,200 or 83% were successfully traced to their new address. The remaining non-response -- that due to causes other than refusal and failure to trace -- is attributed to a wide variety of factors. However, it is likely that a good portion of this is in fact undetected refusals and changes of address. Over time, it is probable that the proportion of non-interviews due to reasons other than refusals and failure to trace will decline.

As noted earlier, SLID interviews people now living with longitudinal respondents who were not present at the beginning of the panel. This is mainly done to ensure that the family information on longitudinal respondents is complete, but it serves a second purpose as well: it improves the capacity of the survey to produce cross-sectional estimates by refreshing the sample. The following table shows a cross-sectional view of response. Because the sample was actually selected in January 1993 (at the beginning of the first reference year), there were already cohabitants identified in January 1994, when the labour and income results for the 1993 reference year were collected. The table shows an eligible population of 32,100 (that is, longitudinal respondents and cohabitants aged 16 and over) in Panel 1 for 1993 and 34,000 one year later. Part of the increase is due to 15 year-olds turning 16, the rest is due to cohabitants:

Panel 1	1993	1994
Eligible population (longitudinal respondents + cohabitants, 16+)	32,100	34,000
Responding	29,400	30,000
Non-response	2,700	4,000
% responding	92%	88%

The results show a substantial amount of change occurring within the family, and this is often associated with geographical mobility. It is a phenomenon that makes the conduct of longitudinal surveys more difficult but, at the same time, increases their usefulness.

V. SUMMARY

The development of the Survey of Labour and Income Dynamics began in 1992. Five years later, we are on the point of releasing the second wave of data. The main design and data collection challenges have been dealt with. The data processing and timeliness concerns are the current focus of our attention. The development of products that are accessible to a wide base of data users and the question of access to detailed microdata are among the challenges ahead.

As for any new survey, there is enthusiasm about SLID among data users. But longitudinal surveys typically take some time to bear fruit, partly because they are complex, but mainly because it requires an accumulation of data in order to create a longitudinal picture. It is important to sustain interest and support for the survey during its early years -- a period of great vulnerability for longitudinal surveys.

With the release of SLID's second wave, it will not yet be possible to study persistence of low income. However, we can examine the magnitude of flows into and out of low income and to look at associated labour market and family events.

The first SLID results on low income dynamics will be published in an environment where there is already extensive use made of cross-sectional income data. The approach will be to expand on the information already in the public domain by adding a new dimension. In time, this information will increase our understanding of the issues affecting the duration of low income spells and the factors that are the most important in triggering flows into and out of poverty. Thus, the data on income dynamics will be complementary to the trend information already available on income levels and income distribution.

APPENDIX

Statistics Canada's Low Income Cut-offs (LICOs)

The Low Income Cut-offs published by Statistics Canada have been described as a hybrid approach, containing elements of both absolute and relative low income measurement. LICOs are calculated using data from the Family Expenditure Survey. This survey, which has typically been done on a national scale every four years or so, is a comprehensive survey of all expenditures made by the household over a calendar year. The survey also collects income data.

The procedure involves determining the overall proportion of income that is spent on food, shelter and clothing, then adding 20 percentage points to that figure. This is an arbitrary amount, but it has been used since the first low income cut-offs were calculated in 1959. The resulting proportion has shifted down from about 70% in 1959 to 56% based on expenditure patterns in 1992.

The Family Expenditure Survey data is then used to estimate income levels by family size and size of community where average expenditure on food, shelter and clothing is 56% of income. There are seven family size categories and five size of community variables, so a table of low income cut-offs has 35 values. With every new Family Expenditure Survey, the LICOs are updated to reflect the current proportion of income spent on food, shelter and clothing. In the intervening years, the Consumer Price Index is used to update the values.