Catalogue no. 75F0002MIE — No. 008
ISSN: 1707-2840
ISBN: 0-662-37553-X

Research Paper

Income research paper series

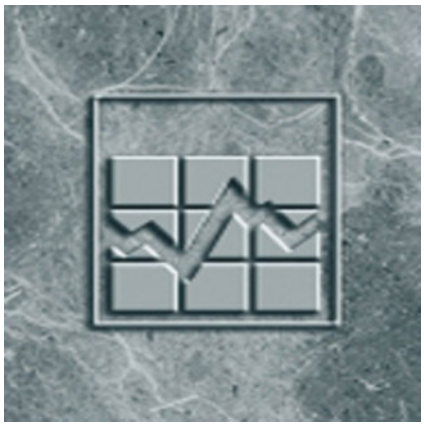
Combined-panel longitudinal weighting Survey of Labour and Income Dynamics

1996-2002

by Jean-François Naud

Income Statistics Division
Jean Talon Building, Ottawa, K1A 0T6

Telephone: 613 951-7355



This paper represents the views of the author and does not necessarily reflect the opinions of Statistics Canada.

Statistics Canada Statistique Canada

Canada

How to obtain more information

Specific inquiries about this product and related statistics or services should be directed to Client Services, Income Statistics Division, Statistics Canada, Ottawa, Ontario, K1A 0T6 ((613) 951-7355; (888) 297-7355; income@statcan.ca).

For information on the wide range of data available from Statistics Canada, you can contact us by calling one of our toll-free numbers. You can also contact us by e-mail or by visiting our Web site.

National inquiries line	**1 800 263-1136**
National telecommunications device for the hearing impaired	**1 800 363-7629**
Depository Services Program inquiries	**1 800 700-1033**
Fax line for Depository Services Program	**1 800 889-9734**
E-mail inquiries	**infostats@statcan.ca**
Web site	**www.statcan.ca**

Ordering and subscription information

This product, Catalogue no. 75F0002MIE2004008, is available on Internet free. Users can obtain single issues at: http://www.statcan.ca/cgi-bin/downpub/research.cgi.

Standards of service to the public

Statistics Canada is committed to serving its clients in a prompt, reliable and courteous manner and in the official language of their choice. To this end, the Agency has developed standards of service which its employees observe in serving its clients. To obtain a copy of these service standards, please contact Statistics Canada toll free at 1 800 263-1136.

Statistics Canada
Income Statistics Division

Income research paper series

Combined-panel longitudinal weighting Survey of Labour and Income Dynamics

1996-2002

Published by authority of the Minister responsible for Statistics Canada

June 2004

Catalogue no. 75F0002MIE2004008

ISSN: 1707-2840

ISBN: 0-662-37553-X

Frequency: Occasional

Ottawa

La version française de cette publication est disponible sur demande (n° 75F0002MIF au catalogue).

Note of appreciation

Canada owes the success of its statistical system to a long-standing partnership between Statistics Canada, the citizens of Canada, its businesses, governments and other institutions. Accurate and timely statistical information could not be produced without their continued cooperation and goodwill.

Summary

The Survey of Labour and Income Dynamics (SLID) is a longitudinal survey composed of panels of six years in length. Since the introduction of the second panel in reference year 1996, two panels overlap for periods of three years. Since the beginning of the survey, two types of weights have been produced for each reference year : a longitudinal weight for each panel and a cross-sectional weight which combines data from both panels. The longitudinal weight for one panel allows conducting analyses relating to the population at the time of its selection and that can be carried out over a period of up to six years. However, some SLID data users have expressed the desire to be able to conduct longitudinal analyses using both panels, and thus increasing their precision. The combined panel longitudinal weight has been created to meet this need. It allows doing analyses which refer to the population at the time of the selection of the most recent panel, using individuals from both panels. However, the analyses are limited to the period of three years during which the two panels overlap. This document presents the principles behind the combined panel longitudinal weighting methodology as well as the steps leading to the creation of the weights. These steps are largely inspired from the steps used in the longitudinal weighting of one panel and the cross-sectional weighting. The results obtained with the new weight are briefly evaluated.

Table of Contents

1. Introduction

The Survey of Labour and Income Dynamics (SLID) is a longitudinal panel survey of individuals. Its goal is to measure changes in the economic well-being of individuals and the factors that influence those changes. When it was introduced in the 1993 reference year, the survey was intended to provide longitudinal data, but over the years, its cross-sectional dimension has become just as important. It uses a household sample composed of two panels that are six years in length and overlap for three years.

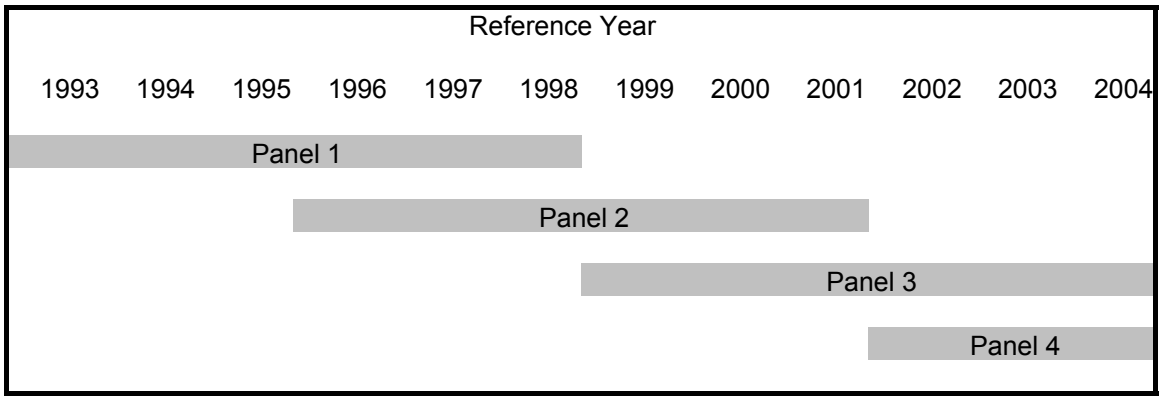
Since the survey's inception, two main types of weights have been produced: a longitudinal weight for each panel, representing the population at the time of selection; and a cross-sectional weight, combining the individuals from both panels for a particular reference year. When longitudinal analyses were conducted, the data for a single panel could be used. Combined-panel longitudinal weighting (CPLW) was developed so that longitudinal studies could use two panels at the same time, thereby doubling the sample size and increasing the precision of the estimates. On the other hand, analyses based on this weighting scheme will be limited to a period of three years.

This paper describes the methodology developed to design and produce combined-panel longitudinal weights. First, SLID's general methodology will be outlined, and a brief description of the longitudinal weighting used for each panel and of the cross-sectional weighting will be provided. Then the various aspects of combined-panel longitudinal weights and the procedure for producing them will be presented. Lastly, there will be a brief assessment of the results obtained with the new weight.

2. SLID Methodology

The Survey of Labour and Income Dynamics is a continuing survey. It is composed of two rotating panels, each six years in length. After the second panel is introduced, there are always two panels at the same time, with each pair of successive panels overlapping for a period of three years. Panel 1 was selected on December 31, 1992, and Panel 2 on December 31, 1995. Since then, a new panel has been selected every three years to replace the older of the two panels, as shown in Figure 1.

Figure 1
Overlap of SLID Panels



The SLID sample consists of about 15,000 households (roughly 40,000 people) for Panel 1 and 17,000 for subsequent panels. The sample is taken from the Labour Force Survey (LFS), whose methodology is described in Singh *et al*. (1990) and Gambino *et al*. (1998). The LFS operates on the basis of six panels; each panel remains in the sample for six months, with one panel replaced each month. The last-stage sampling unit is the dwelling. All members of the households occupying the selected dwellings are included in the LFS sample.

The sample for a SLID panel is composed of households from two outgoing LFS rotation groups in January and February of the first reference year. SLID selects only households that were LFS respondents in January. The final LFS interview serves as the introductory contact for SLID (Lavigne and Michaud, 1998).

Hence, the initial weight for SLID comes directly from LFS and is at the household level. It is used to compute the longitudinal and cross-sectional weights for each wave.

3. Current Longitudinal and Cross-sectional Weights

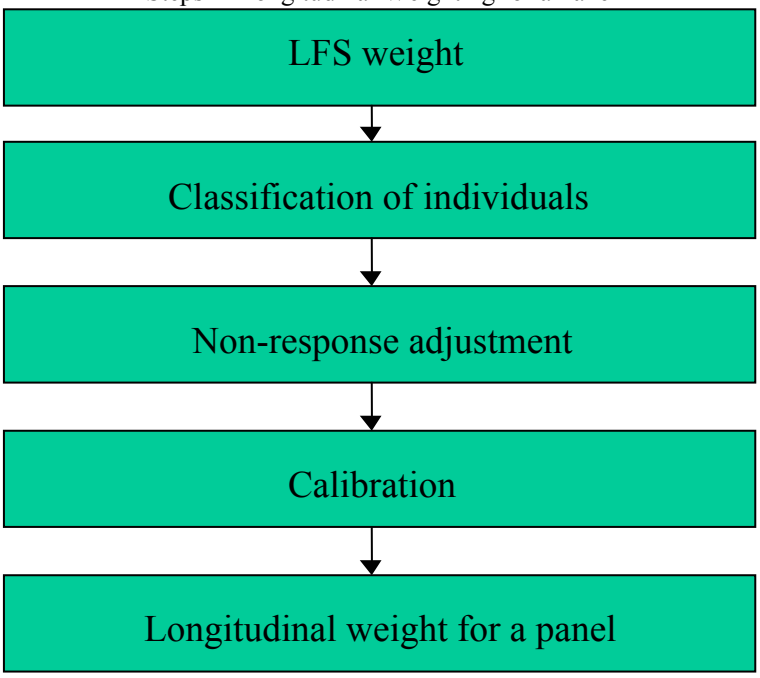
Since it was introduced in 1993, SLID has produced two main types of weights for each reference year: longitudinal weights for each panel and cross-sectional weights for the two panels combined. Combined-panel longitudinal weighting borrows elements from both methodologies. This section contains a brief description of the survey's longitudinal and cross-sectional weights to help the reader understand the development of the new methodology presented in section 4.

3.1 Longitudinal weighting

Before the introduction of combined-panel longitudinal weighting, SLID produced just one type of longitudinal weight. That weight is specific to one of the current panels and represents the population at the time of its selection. It can be used to conduct studies covering the panel's entire six-year lifespan. An overview of the methodology used to produce a panel's longitudinal weight is presented here to make combined-panel longitudinal weighting easier to understand. For a detailed description, see Lévesque and Franklin (2000). In addition, since the steps are also part of the combined-panel longitudinal weighting methodology, they will be described at greater length in section 4.

The target population associated with the longitudinal weight is the population at the time of the panel's selection (December 31, 1992, for Panel 1; December 31, 1995, for Panel 2; and so on). The sample consists of all members of the selected households at the beginning of the panel (longitudinal) and excludes people who joined the households subsequently (cohabitants). The initial weight used for longitudinal weighting is an LFS household weight. However, since the basic longitudinal unit is the individual, SLID's longitudinal weight is at the individual level.

Figure 2
Steps in Longitudinal Weighting for a Panel



Several steps are needed to derive the longitudinal weight for a panel (Figure 2). First, individuals are classified according to whether they are respondents, non-respondents or out of scope (deceased, institutionalized, or outside the 10 provinces). Respondents and out-of-scope individuals will have a non-zero longitudinal weight, and non-respondents will have a weight of zero.

The next step is non-response adjustment. A non-response model is developed, and the weights of respondents are adjusted so that they represent non-respondents as well. Out-of-scope individuals retain their initial weight, thereby representing the portion of the target population that was present at the time of the panel's selection and subsequently left the 10 provinces, entered an institution or died.

Next, calibration is performed to ensure that certain totals computed with the weights match the population totals derived from other sources. Those totals are, for each province, the number of individuals in each age-sex group, the number of size 1 and 2 economic families, and the number of size 1 and 2 households. They apply to the longitudinal target population, i.e., the population at the time the panel was selected. The result is the final longitudinal weight for the panel. That weight is produced for each reference year. Note that in the near future, calibration will also be based on salary classes (Latouche and Laroche, 2003).

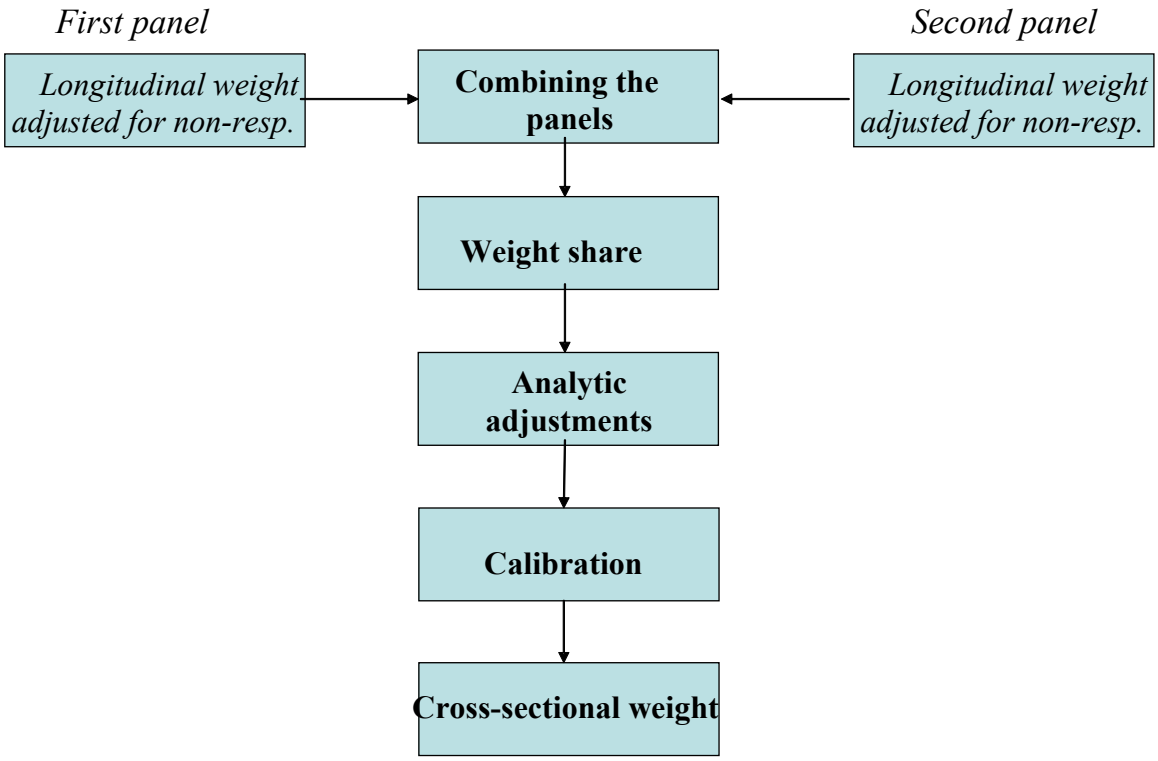
3.2 Cross-sectional weighting

The SLID cross-sectional weight is used to produce estimates for a particular reference year. To that end, the two panels are combined. Individuals who have joined the households of longitudinal persons are referred to as cohabitants and are also part of the cross-sectional sample. A brief overview of the cross-sectional weighting methodology will be presented here. For additional information, see Lévesque and Franklin (2000).

The cross-sectional weight's target population is the population of the 10 provinces on December 31 of the reference year, excluding people living on reserves, in institutions or in military barracks. All longitudinal persons and individuals living in their households (cohabitants) are part of the cross-sectional sample. The

initial cross-sectional weight is the longitudinal weight adjusted for non-response, which represents the population at the time each panel was selected.

Figure 3
Steps in the Cross-sectional Weighting Process



The first step in the cross-sectional weighting process is to combine the samples for the two panels by applying an allocation factor to the non-response-adjusted longitudinal weight. The allocation factor is computed – separately for each province – so as to minimize the variance of a point estimate. No adjustment factor is applied to the weights of Panel 2 individuals who could not have been selected for Panel 1. The panel combination step in the combined-panel longitudinal weighting process, described in section 4.9, is very similar. For more details on panel allocation factors, see Latouche *et al*. (2000) and Merkouris (1999).

The next step in the cross-sectional weighting process is the weight share (Lavallée, 1995). It transfers part of the weight from longitudinal persons to cohabitants who joined their households. Then come the analytic adjustments, one for interprovincial migration and the other for influential values. Lastly, as in the case of longitudinal weighting, calibration is performed against known totals for the reference year (the number of individuals in each age-sex group, the number of size 1 and 2 economic families, and the number of size 1 and 2 households). The result is the final cross-sectional weight. It is produced for each reference year. Note that as in the case of longitudinal weighting, calibration will also be based on salary classes in the near future (Latouche and Laroche, 2003).

4. Combined-Panel Longitudinal Weighting

The purpose of combined-panel longitudinal weighting is to make it possible for longitudinal analyses to use the samples from both panels and thus benefit from the extra precision gained by doubling the sample size. However, since the two panels overlap for only three years, the analyses cannot cover longer periods. Like the other types of weights, a combined-panel longitudinal weight is computed for each reference year.

This section describes in detail the methodology of combined-panel longitudinal weighting. First, the rationale for creating this new weight for SLID will be presented, along with the weight's limitations. This will be followed by information about the target population and the sample. Finally, the steps involved in generating the weight will be outlined.

4.1 The issues

Previously, SLID had two main types of weights: a longitudinal weight for each panel and a cross-sectional weight for the two current panels combined. Longitudinal analyses were, of necessity, based on the sample in just one panel, about 40,000 people. A number of users expressed interest in longitudinal analyses that would be based on both panels. Combining the two panels doubles the sample size and increases the precision of the estimates.

However, since the panels overlap for three years, it is difficult to perform longitudinal studies combining the two panels over a long period. For example, with the weight for Panels 1 and 2, studies can cover only the period from December 31, 1995, to December 31, 1998 (in the case of the weight produced for the 1998 reference year), i.e., from the start of Panel 2 to the end of Panel 1.

4.2 Definitions

To make the text easier to read and understand, the following terms will be used:

First panel:	Denotes the older of the two panels being combined. For example, in the longitudinal weighting of Panels 1 and 2 combined, the first panel will be Panel 1. When Panels 2 and 3 are combined, it will be Panel 2.
Second panel:	Denotes the younger of the two panels being combined. For example, in the longitudinal weighting of Panels 1 and 2 combined, the second panel will be Panel 2. When Panels 2 and 3 are combined, it will be Panel 3.
Panel combination date:	The date associated with the target population of the combined-panel longitudinal weight. The date is also associated with the second panel's target population. For the combination of Panels 1 and 2, the date is December 31, 1995. For Panels 2 and 3, it is December 31, 1998.

In the rest of this paper, the terms "weighting" or "weight" without further qualification are to be understood as referring to combined-panel longitudinal weighting.

4.3 Target population

The target populations of the two panels being combined are different. In both cases, the target population is the Canadian population living in one of the 10 provinces, excluding Indian reserves, military barracks and institutions. However, the two populations are separated by three years. Since the panels begin overlapping as soon as the second panel is introduced, the target population must be identical to the second panel's longitudinal population. For example, when Panels 1 and 2 are combined, the target population will be the same as the Panel 2 target population, i.e., the target population on December 31, 1995.

4.4 Sample used

The combined-panel longitudinal weight must be representative of the target population, and therefore only units that are in that population can be kept in the sample. First-panel longitudinal individuals not in the target population (on the date on which the panels were combined) must be removed from the combined longitudinal sample, even if they may end up in the target population for subsequent years. Those which *are* in the target population will be included in the sample, even if they drop out of the target population in a subsequent year. But it is impossible to identify individuals who are not in the target population if they were non-respondents for the year in which the panels began overlapping. As a result, they have to be considered part of the target population, and the sample, even if they should theoretically be excluded from it. All second-panel longitudinal individuals are part of the target population and the sample, whether they are respondents or not. This does not mean that they will all have a weight greater than zero: non-respondents will have a weight of zero.

For example, when Panels 1 and 2 are combined, the combined longitudinal sample will consist of Panel 1 individuals still living in one of the 10 provinces (excluding institutions, reserves and military barracks) on December 31, 1995, and of all Panel 2 individuals. Panel 1 individuals who were non-respondents in 1995 will have to be included in the target population.

Figure 4
Sample for Combined-Panel Longitudinal Weighting
Example of combining Panels 1 and 2, 1998 reference year

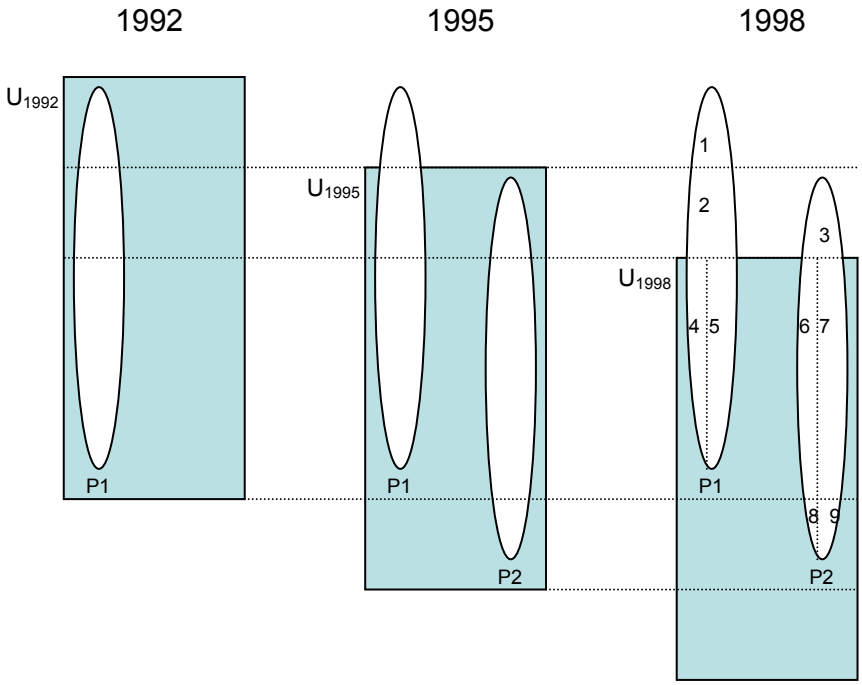


Figure 4 shows the sample used for combining Panels 1 and 2 for the purpose of analyses covering the period from 1995 to 1998 (with the weight produced for the 1998 reference year).

U_{1992} : SLID target population on December 31, 1992 (selection of first panel)
U_{1995} : SLID target population on December 31, 1995 (selection of second panel and panel combination date)
U_{1998} : SLID target population on December 31, 1998 (reference date for the current year)
P1 : Panel 1 sample
P2 : Panel 2 sample

The figures in the various zones of the 1998 samples relate to different categories of individuals:

1: Panel 1 individuals who are not in the target population for combined-panel longitudinal weighting (i.e., on December 31, 1995). These individuals have a weight of zero. Note that the chart has been simplified; in theory, a Panel 1 individual could be out of scope in 1995 and back in the target population in 1998. In that case, he/she would also have a weight of zero.

2 and 3: Individuals from the two panels who are in the target population for combined-panel longitudinal weighting but not in the target population for the 1998 survey. Their weight is greater than zero.

4, 6 and 8: Individuals from the two panels who are in the target population for combined-panel longitudinal weighting (on December 31, 1995) and the target population for the current-year survey (on December 31, 1998) but are non-respondents. These individuals have a weight of zero.

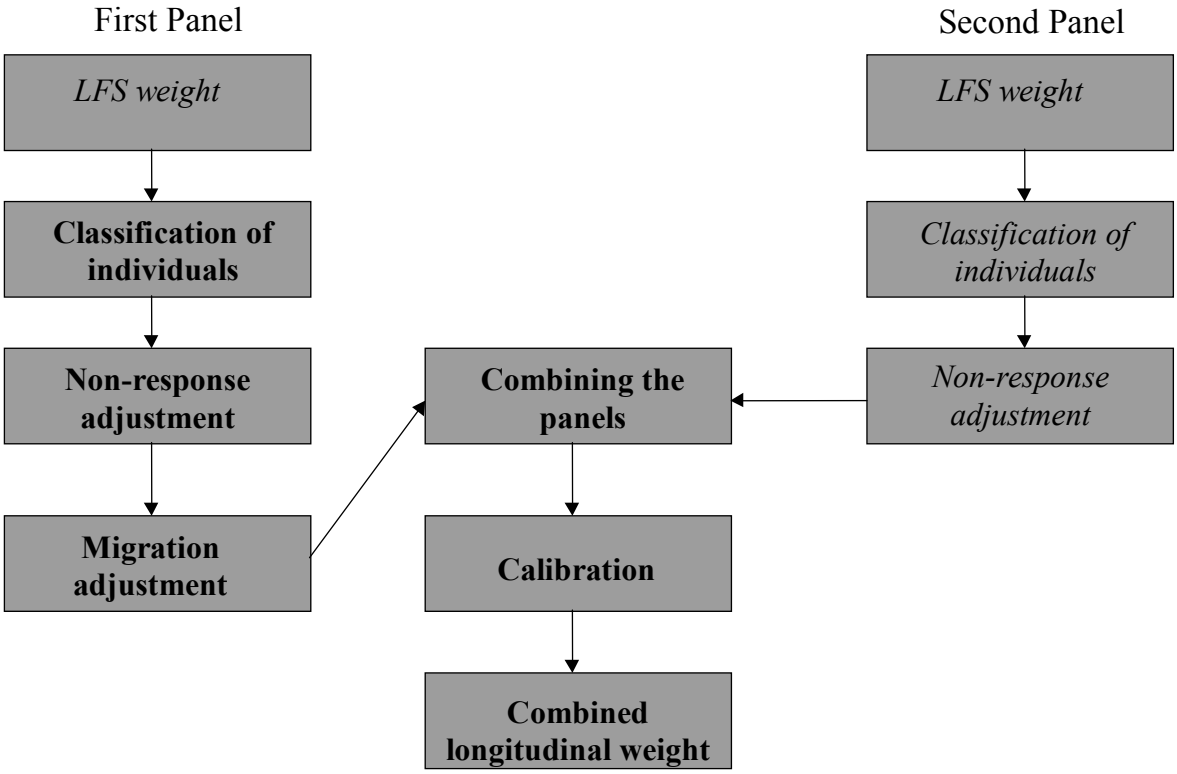
5, 7 and 9: Individuals from the two panels who are in the target population for combined-panel longitudinal weighting (on December 31, 1995) and the target population for the current-year survey (on December 31, 1998) and are respondents. These individuals have a weight greater than zero.

Hence, the sample consists of all individuals from the two panels except the ones in zone 1. Individuals in zones 2, 3, 5, 7 and 9 will have a weight greater than zero.

4.5 Steps in the combined-panel longitudinal weighting process

Producing the combined-panel longitudinal weight is very similar to producing the longitudinal weight for one of the panels. However, the process is more complicated because it involves combining the samples for two panels whose target populations are three years apart.

Figure 5
Steps in the Combined-Panel Longitudinal Weighting Process



Since the two samples were selected independently on dates three years apart, they have to be treated separately at the beginning of the process.

For the second panel, since it already represents the target population on the combination date, the classification of individuals and non-response adjustment will be exactly the same as for the regular longitudinal weighting. Consequently, it will not be necessary to redo these steps, and the panel's non-response-adjusted weight can be used directly in the panel combination step.

For the first panel, the initial steps have to be redone because the individuals who are not in the target population on panel combination date must be removed from the sample. In other words, the classification of individuals and the non-response adjustment must be redone using only those individuals who are in the target population for combined-panel longitudinal weighting. Then an adjustment is made for interprovincial migration to ensure that individuals who moved from one province to another between the

first panel's sample selection date and the panel combination date do not have a weight that is excessively large relative to other weights.

Next, the panels are combined to produce a weight for all individuals that represents the target population. The last step, as in the case of longitudinal weighting for each panel, is to perform calibration to ensure that the estimates computed with the weights match certain known totals for the population.

4.6 Classification of individuals

As in the case of longitudinal weighting for each panel, the first step in combined-panel longitudinal weighting is the classification of individuals. The two panels are dealt with separately.

For the second panel, the classification will be the same as it was for regular longitudinal weighting. Individuals will be placed in different categories based on whether they are respondents, non-respondents or cross-sectionally out of scope for the reference year.

For the first panel, the process will be identical in all respects except one: individuals who are not in the target population on panel combination date are removed from the sample. In addition, persons who did not respond to the survey for the year in which the second panel was introduced are considered non-respondents for the current reference year, even if they actually responded that year. This is necessary because certain characteristics of respondents on panel combination date are needed for some of the subsequent steps in the weighting process, including calibration.

Table 1
Classification of Individuals
Example: the weighting of Panels 1 and 2 combined, 1998

Categories (from Figure 4)	Description	Number of individuals
Panel 1		
1	Not in the target population	980
2	Out of scope (cross-sectionally)	1,110
4	Non-respondents	7,493
5	Respondents	30,032
Total		39,615
Panel 2		
3	Out of scope (cross-sectionally)	1,244
6 and 8	Non-respondents	6,458
7 and 9	Respondents	35,842
Total		43,544
Panels 1 and 2 combined		
Total		83,159

Table 1 shows the counts for the 1998 reference year. Since about 1,000 individuals from the first panel are not in the target population, the second panel will obviously be more important. Furthermore, because of sample erosion and the slightly more restrictive definition of respondent for the first panel, the number of respondents is higher for the second panel.

4.7 Non-response adjustment

After respondents and non-respondents have been identified, the next step is to adjust for non-response. The only data available for both respondents and non-respondents are the data from the preliminary interview (Panel 1) or the final LFS interview (Panels 2, 3 and 4). Those data are used for modelling. Since the data used for first-panel individuals are three years older than the data used for second-panel

individuals, non-response must be modelled separately for each panel. For Panel 1, for example, even though the target population is the population on December 31, 1995, the adjustment must be based on the data from the preliminary interview (December 31, 1992).

As in the case of longitudinal weighting for each panel, response homogeneity groups (RHGs) are generated by segmentation modelling using chi-square automatic interaction detection (CHAID) (see Kass, 1980), which selects the variable with the highest Pearson chi-square statistic. The weights are then adjusted for non-response in each RHG. The non-response-adjusted weights for the second panel are exactly the same as when the panel is weighted on its own. For the first panel, the adjustment must be redone since individuals who were not in the target population on the panel combination date have been removed from the sample and since the definition of a respondent has an added condition: he/she must also have been a respondent in the year in which the second panel was introduced.

For more details on the SLID non-response adjustment method, see Lévesque and Franklin (2000).

4.8 Migration adjustment

The initial longitudinal weight of first-panel individuals is representative of the target population at the time of its selection, and therefore the weight relates to the province at the time of selection and not to the province of residence on combination date. If in the meantime the individual has moved from a province with a very small sampling fraction (high weight) to a province with a large sampling fraction (low weight), he/she may end up having far too much influence on the new province's estimates.

An interprovincial migration adjustment is made in combined-panel longitudinal weighting for units from the first panel.[1] This is primarily due to the fact that we want in-scope individuals from the first panel to have a longitudinal weight that is representative of the new target population, which is the population at the time the second panel is selected (or on panel combination date).

The adjustment lowers the weight of people who move to a particular province if it is higher than the maximum weight of non-migrants of that same province. The new weight is equal to the 95^{th} percentile in the distribution of non-migrants' weights.

4.9 Combining the panels

At this point, a non-response-adjusted longitudinal weight is available for each panel. For the first panel, that weight is representative of the target population excluding the people who joined it at some point in the three years separating the two panels. For the second panel, the weight already represents the target population. The next step, then, is to combine the two panels to obtain a single panel containing individuals who represent the target population on panel combination date.

Since all first-panel longitudinal individuals who are in scope on panel combination date (zones 2, 4 and 5 in Figure 4) can also be selected for the second panel, their weight is multiplied by a factor p_1 between 0 and 1. The weight of second-panel longitudinal individuals who could have been selected for the first panel (zones 3, 6 and 7 in Figure 4) is multiplied by a factor $p_2 = 1 - p_1$. No factor is applied to second-panel longitudinal persons who were not in the target population when the first panel was selected (zones 8 and 9 in Figure 4). Those individuals were not yet born, outside the 10 provinces, institutionalized or members of the Armed Forces living in barracks when the first panel was selected. Only newborns and international immigrants can be identified. The rest will be considered the same as second-panel longitudinal individuals who could have been selected for the first panel, and their weight will be multiplied by p_2.

1. A similar adjustment is made in SLID's cross-sectional weighting.

The formula used to calculate the panel allocation factors for combined-panel longitudinal weighting is the same as the one used for cross-sectional weighting (Latouche *et al.*, 2000). It minimizes the variance of point estimates made with the sample, before calibration, for all reference years, including the target population (on panel combination date).

Calculation of the panel allocation factors (p) is based on the following formula:

$$p_1 = \frac{n_1}{n_1 + n_2(\frac{d_1}{d_2})}$$

$$p_2 = 1 - p_1$$

An allocation factor is computed for each province. The variables n_1 and n_2 represent the number of individuals aged 16 and over in the two panels who will have a combined longitudinal weight. Only individuals aged 16 and over are considered because there are no income or labour force activity data for children. The variables d_1 and d_2 represent the design effect for the two panels. The design effect is defined as the ratio of the variance obtained with the survey's design to the variance that would be obtained with simple random sampling. The design effect used is the LFS design effect (SLID is an LFS supplement) at the time of the panel's selection; the LFS design effect is associated with the estimated number of people aged 16 and over in the province.

We also considered using allocation factors that minimize the variance of a trend between waves t and t+1. However, that would require more detailed studies, and the increase in precision would probably not be very large in relation to the method used. The following formula is provided for reference purposes:

$$p_1 = \frac{V(\hat{Y}_{t+1,2}) + V(\hat{Y}_{t,2}) - 2COV(\hat{Y}_{t+1,2}, \hat{Y}_{t,2})}{\sum_{j=1}^{2}[V(\hat{Y}_{t+1,j}) + V(\hat{Y}_{t,j}) - 2COV(\hat{Y}_{t+1,j}, \hat{Y}_{t,j})]}$$

where $\hat{Y}_{t,p}$ is the estimate, based on panel p, of a total or average for wave t.

4.10 Adjustment for influential values

The income distribution is asymmetric, with a very long tail for the higher values. As a result, one or more individuals who have both a very large income and a high weight may have an excessive influence on the average income estimates for provinces or smaller domains, and on the associated variance estimate. For that reason, an adjustment for influential values was incorporated in SLID's weighting strategy to lower the weight of such individuals and reduce their influence.

The method used, developed by Tremblay (1998), is applied cross-sectionally. The adjustments are subsequently used for longitudinal weighting as well. The cross-sectional weighting methodology is explained in detail in Lévesque and Franklin (2000).

The influential-value adjustments computed for cross-sectional weighting are applied unchanged during combined-panel longitudinal weighting. However, since this involves longitudinal weighting, the adjustments computed for the preceding years are also applied to the individual's weight. If the individual received an adjustment for more than one reference year, the largest adjustment (i.e., the one with the smallest factor) is applied.

4.11 Calibration

The final step in producing the combined-panel longitudinal weight is calibration on margins. As in the case of longitudinal weighting for each panel, we want the sum of the weights to be equal to certain totals that are known for the target population on panel combination date. Calibration is performed for each province. The control totals used are estimates based on census data: population counts for age-sex groups, the number of size 1 and 2 economic families, and the number of size 1 and 2 households. As in the case of longitudinal weighting for each panel and of cross-sectional weighting, calibration will also be based on salary classes in the near future (Latouche and Laroche, 2003).

The calibration method is generalized regression (GREG).

The resulting weight is the final combined-panel longitudinal weight, which will be used to produce the estimates.

5. Evaluation

Since the combined-panel longitudinal weight always represents the population at the time the panels are combined, it is difficult to evaluate the estimates based on this weight by comparing them with data from other sources. However, since the second panel's longitudinal weight represents the same target population, it is possible to compare the estimates produced with the two weights and the associated variances. The variances are estimated by the bootstrap method (Efron, 1982; Rao and Wu, 1987; Rao, Wu and Yue, 1992). Note that the analysis of the differences between the estimates produced with the two weights can also be interpreted as a comparison of Panels 1 and 2.

Table 2
Place of Residence of Individuals on December 31, 1998

	Estimate Panel 2 longitudinal weight	Estimate Combined-panel longitudinal weight	Percentage difference	Coefficient of variation (%) Panel 2 longitudinal weight	Coefficient of variation (%) Combined-panel longitudinal weight	Significant difference between the estimates Confidence level = 0.05 (*)
Newfoundland	515 615	522 620	1.36	1.36	1.01	*
Prince Edward Island	124 397	125 431	0.83	1.66	1.42	
Nova Scotia	886 304	881 629	-0.53	1.61	1.25	
New Brunswick	707 558	706 316	-0.18	1.92	1.53	
Quebec	6 928 676	6 913 235	-0.22	0.36	0.29	
Ontario	10 521 030	10 518 677	-0.02	0.38	0.32	
Manitoba	1 003 857	1 006 685	0.28	1.06	0.87	
Saskatchewan	929 087	925 341	-0.40	1.31	1.04	
Alberta	2 637 129	2 652 800	0.59	1.19	0.92	
British Columbia	3 598 511	3 599 722	0.03	0.87	0.69	
In an institution	106 295	103 475	-2.65	10.73	9.17	
Deceased	465 291	466 769	0.32	4.69	3.91	
Total (including persons outside the 10 provinces)	28 733 700	28 733 700	0.00	0.00	0.00	

Table 2 shows the distribution of the longitudinal population by place of residence on December 31, 1998, estimated with the Panel 2 longitudinal weight and with the combined-panel longitudinal weight, the percentage difference between the estimates, and the coefficient of variation for each estimate. The last column indicates whether the difference between the estimates is significant at the 0.05 level; significance is determined from an estimate of the variance of the difference, which is not included in the table. The estimates are very similar, as the coefficients of variation show a significant difference only for Newfoundland. Since the weights are calibrated on the December 31, 1995, population counts, it was to be expected that the estimates would be similar at the provincial level and identical for the total. The estimates produced with the combined-panel longitudinal weight are more precise in every case, though the difference in precision is not very large.

Table 3
Average Income by Province, 1998

	Estimate Panel 2 longitudinal weight	Estimate Combined-panel longitudinal weight	Percentage difference	Coefficient of variation (%) Panel 2 longitudinal weight	Coefficient of variation (%) Combined-panel longitudinal weight	Significant difference between the estimates Confidence level = 0.05 (*)
Newfoundland	18 697	18 429	-1.44	4.10	2.56	
Prince Edward Island	20 665	21 230	2.73	3.46	2.79	
Nova Scotia	21 329	21 468	0.65	2.82	2.35	
New Brunswick	21 965	21 681	-1.29	2.90	2.09	
Quebec	23 566	23 661	0.40	1.99	1.55	
Ontario	28 359	28 979	2.19	1.73	1.63	*
Manitoba	24 143	23 892	-1.04	3.04	2.32	
Saskatchewan	23 283	23 515	1.00	3.15	2.40	
Alberta	28 352	28 369	0.06	3.23	2.32	
British Columbia	25 955	26 411	1.75	2.44	2.08	
All 10 provinces combined	25 919	26 232	1.21	1.01	0.87	*

Table 3 is similar to Table 2, but it presents estimates of average income by province. The estimates of 1998 average income differ slightly more than the population estimates. For most provinces, the estimate produced with the combined-panel longitudinal weight is slightly higher, although the difference is either not significant or only marginally so. As in the case of Table 2, the use of the combined-panel longitudinal weight increases the precision in all cases, though the increase is quite modest.

Table 4
Proportion of Individuals Below the Low-Income Cut-off, 1998

	Estimate Panel 2 longitudinal weight	Estimate Combined-panel longitudinal weight	Percentage difference	Coefficient of variation (%) Panel 2 longitudinal weight	Coefficient of variation (%) Combined-panel longitudinal weight	Significant difference between the estimates Confidence level = 0.05 (*)
Newfoundland	13.38%	13.31%	-0.55	12.87	8.58	
Prince Edward Island	9.49%	7.85%	-17.34	20.03	18.67	*
Nova Scotia	14.85%	14.18%	-4.49	9.29	7.82	
New Brunswick	10.03%	10.00%	-0.34	12.98	9.74	
Quebec	16.08%	15.96%	-0.76	7.08	5.49	
Ontario	10.54%	10.03%	-4.83	6.53	5.88	*
Manitoba	14.58%	13.59%	-6.82	11.19	8.93	
Saskatchewan	9.80%	9.29%	-5.20	12.82	10.02	
Alberta	12.10%	11.64%	-3.80	8.15	6.35	
British Columbia	11.77%	11.03%	-6.30	9.08	7.87	
All 10 provinces combined	12.52%	12.07%	-3.58	3.57	2.98	*

Since estimated average income is slightly higher using the combined-panel longitudinal weight, it is no surprise that the estimates of the proportion of individuals below the low-income cut-off would be lower. The differences are appreciable, nearly 0.5% for all provinces combined. The coefficients of variation of the difference also indicate that it is significant for Prince Edward Island, Ontario and all 10 provinces combined. Unfortunately, it is impossible to determine which of the two estimates is more accurate. Nevertheless, the estimates obtained with the combined-panel longitudinal weight are more precise.

Table 5
Average Difference in Total Income Between 1997 and 1998

	Estimate Panel 2 longitudinal weight	Estimate Combined-panel longitudinal weight	Percentage difference	Coefficient of variation (%) Panel 2 longitudinal weight	Coefficient of variation (%) Combined-panel longitudinal weight	Significant difference between the estimates Confidence level = 0.05 (*)
Newfoundland	660	597	-9.49	46.76	33.86	
Prince Edward Island	994	1 048	5.42	31.66	26.96	
Nova Scotia	1 056	1 064	0.76	20.58	17.22	
New Brunswick	1 257	1 238	-1.52	16.37	13.00	
Quebec	1 350	1 338	-0.87	14.08	11.57	
Ontario	1 391	1 706	22.63	14.39	15.17	
Manitoba	1 614	1 531	-5.10	19.65	14.99	
Saskatchewan	874	985	12.76	33.09	21.99	
Alberta	1 680	1 743	3.78	35.94	23.18	
British Columbia	851	804	-5.58	30.94	26.41	
All 10 provinces combined	1 299	1 414	8.83	8.50	8.05	

Table 5 presents estimates of the average increase in personal total income between 1997 and 1998. The large percentage differences between the estimates based on the two types of weights are due to the nature of the variable being estimated. The coefficients of variation are very high, especially for the difference between the estimates. The precision obtained with the combined-panel longitudinal weight is not much better than the precision obtained with the Panel 2 longitudinal weight.

Table 6
Number of Years Below the Low-Income Cut-off Between 1996 and 1998

	Estimate Panel 2 longitudinal weight	Estimate Combined-panel longitudinal weight	Percentage difference	Coefficient of variation (%) Panel 2 longitudinal weight	Coefficient of variation (%) Combined-panel longitudinal weight	Significant difference between the estimates Confidence level = 0.05 (*)
0	22 101 268	22 307 533	0.93	0.65	0.53	*
1	2 341 248	2 263 194	-3.33	3.78	3.21	*
2	1 526 475	1 465 215	-4.01	5.43	4.76	*
3	1 883 172	1 816 514	-3.54	5.31	4.44	*
Total	27 852 163	27 852 456	0.00	0.15	0.13	

Table 6 shows the distribution of individuals by the number of years they have spent below the low-income cut-off between 1996 and 1998, the longest possible period for an analysis based on the combined-panel longitudinal weight. The study includes only individuals for whom data are available for all three years. The differences between the estimates produced with the two weights are consistent with the observations regarding previous tables and are significant. The estimates produced with the combined-panel longitudinal weight are slightly more precise.

We might expect the increase in precision yielded by combined-panel longitudinal weighting to be larger for statistics computed for smaller domains. To test this hypothesis, we produced two sets of estimates: the proportion of lone-parent families living below the low-income cut-off, and the average income of immigrants.

Table 7
Proportion of Lone-Parent Families Below the Low-Income Cut-off, 1998

	Estimate Panel 2 longitudinal weight	Estimate Combined-panel longitudinal weight	Percentage difference	Coefficient of variation (%) Panel 2 longitudinal weight	Coefficient of variation (%) Combined-panel longitudinal weight
Newfoundland	43.22%	44.66%	3.33	30.51	20.39
Prince Edward Island	32.82%	29.29%	-10.74	33.91	30.86
Nova Scotia	70.28%	63.36%	-9.84	9.18	8.59
New Brunswick	43.23%	42.93%	-0.68	14.46	11.82
Quebec	40.05%	39.43%	-1.55	13.93	11.71
Ontario	42.76%	42.23%	-1.24	8.76	7.56
Manitoba	48.43%	47.19%	-2.56	24.03	16.14
Saskatchewan	21.36%	20.67%	-3.24	65.48	47.14
Alberta	31.94%	33.95%	6.27	22.20	15.34
British Columbia	48.29%	39.86%	-17.45	16.55	15.79
All 10 provinces combined	41.86%	40.49%	-3.29	6.01	5.16

Table 7 presents provincial estimates of the proportion of lone-parent families living below the low-income cut-off. Although the estimates are for relatively small domains, the percentage differences between the estimates produced with the two weights are, in most cases, similar to the differences in Table 4. The coefficients of variation are higher, though. It is also clear, especially at the provincial level, that combined-panel longitudinal weighting provides an appreciable increase in precision.

Table 8
Average Income of Immigrants, 1998

	Estimate Panel 2 longitudinal weight	Estimate Combined-panel longitudinal weight	Percentage difference	Coefficient of variation (%) Panel 2 longitudinal weight	Coefficient of variation (%) Combined-panel longitudinal weight
Newfoundland	44 409	40 146	-9.60	38.76	22.08
Prince Edward Island	23 848	26 401	10.70	18.34	11.37
Nova Scotia	23 129	25 015	8.15	14.18	10.87
New Brunswick	32 301	29 365	-9.09	12.37	9.69
Quebec	18 723	19 215	2.63	7.15	5.93
Ontario	27 166	28 043	3.23	2.81	2.59
Manitoba	25 363	23 673	-6.66	6.15	5.01
Saskatchewan	27 325	29 020	6.20	22.17	14.33
Alberta	28 577	28 370	-0.73	7.55	5.47
British Columbia	24 024	24 499	1.98	5.88	4.72
All 10 provinces combined	25 513	26 108	2.33	2.24	1.95

Table 8 presents the average personal income of immigrants and hence is similar to Table 3 except that it covers much smaller domains. The percentage difference between the estimates produced with the two weights is larger for the average income of immigrants than for the average income of all individuals. The increase in precision obtained with the combined-panel longitudinal weight is substantial at the provincial level.

The foregoing tables show that there are differences between the estimates produced with the Panel 2 longitudinal weight and the combined-panel longitudinal weight, even though they are associated with the same target population (the population of the 10 provinces on December 31, 1995) and the samples are partly composed of the same individuals. However, when the sizes of the coefficients of variation are taken into account, the differences are small. They suggest that globally, there is a difference between Panels 1 and 2. However, it is expected that calibration on salary classes, which will be incorporated into the survey's weighting strategy in the near future, will reduce the differences.

Combined-panel longitudinal weighting increases the precision of the estimates in almost every case, but the increase is generally quite modest. On the other hand, most of the estimates presented in the above tables apply to large domains. For smaller subpopulations, there is a much larger gain in precision. Moreover, because Panel 1 is more variable, with rather low allocation factors (between 0.18 and 0.43), the extent to which the combined-panel longitudinal weight can increase the precision relative to the Panel 2 weight is limited. When Panels 2 and 3, which have allocation factors of about 0.5, are combined, the increases in precision are more substantial.

The combined-panel longitudinal weighting methodology described here, including some of the findings in this section, was presented to Statistics Canada's Advisory Committee on Statistical Methods in November 2001 (Latouche, 2001).

6. Conclusion

The combined-panel longitudinal weight was created to provide SLID analysts with the ability to perform longitudinal studies based on two panels. Such studies are more precise, but the longest possible period they can cover is three years, which is the length of the overlap between two successive panels.

The combined-panel methodology is based on SLID's current longitudinal and cross-sectional weighting methodologies. It adjusts the sample of the older panel and combines it with the younger sample to produce a new sample representing the target population. For the older panel, the classification of individuals and the non-response adjustment must be redone. An interprovincial migration adjustment is also performed to ensure that individuals who moved from one province to another during the three-year overlap between the two panels do not have an excessive weight. Then the panels are combined in much the same way as in cross-sectional weighting. Finally, as in the case of longitudinal weighting for each panel, an adjustment is made for influential values and calibration is performed.

To evaluate the estimates produced with the new weight, we compared a number of them for the 1998 reference year with estimates produced with the Panel 2 longitudinal weight, which represents the same target population. We found that for large domains, there are some differences in the estimates, but only a few are significant. There is also an increase in precision with the combined-panel weight. The increase in precision is larger for estimates based on smaller domains.

References

Efron, B. (1982), *The jacknife, the bootstrap and other resampling plans*, Philadelphia: Society for industrial and applied mathematics.

Gambino, J. G., Singh, M. P., Dufour, J., Kennedy, B. and Lindeyer, J. (1998). *Methodology of the Canadian Labour Force Survey,* Statistics Canada. Catalogue 71-526-XPB.

Kass, G.V. (1980). *An exploratory technique for investigating large quantities of categorical data.* Applied Statistics, 29, 119-127.

Latouche, M. (2001). *The Survey of Labour and Income Dynamics: Development of a Combined Panel Longitudinal Weight.* Technical report presented to the Advisory Committee on Statistical Methods, November 2001. Statistics Canada. Internal document.

Latouche, M., Dufour, J. and Merkouris, T. (2000). *Cross-sectional Weighting : Combining Two or More Panels*, Statistics Canada. Catalogue 75F0002MIE-00006.

Latouche, M. and LaRoche, S. (2003), *The Survey of Labour and Income Dynamics Calibration Strategy*, Technical report presented to the Advisory Committee on Statistical Methods, November 2003. Statistics Canada. Internal document.

Lavallée, P. (1995). *Cross-sectional Weighting of the Longitudinal Surveys of Individuals and Households Using the Weight Share Method,* Survey Methodology, vol. 21, no 1, pp. 25-32.

Lavigne, M. and Michaud, S. (1998). *General Aspects of the Survey of Labour and Income Dynamics,* Statistics Canada. Catalogue 75F002MIE-1998005.

Lévesque, I. and Franklin, S. (2000). *Longitudinal and Cross-sectional Weighting of the Survey of Labour and Income Dynamics, 1997 Reference Year,* Statistics Canada. Catalogue 75F0002MIE-00004.

Merkouris, T. (1999). *Cross-sectional Estimation in Multiple Panel Household Surveys*. Statistics Canada. Working paper HSMD 99-004.

Rao, J.N.K. and Wu, C.F.J. (1987). *Resampling Inference with Complex Survey Data*, Journal of the American Statistical Association, 83, 231-241.

Rao, J.N.K., Wu, C.F.J. and Yue, K. (1992). *Some Recent Work on Resampling Methods for Complex Surveys*, Survey Methodology, vol.18, n^{o}2, 209-217.

Singh, M. P., Drew, J. D., Gambino, J. G. and Mayda, F. (1990). *Methodology of the Canadian Labour Force Survey*, Statistics Canada. Catalogue 71-526.

Tremblay, J. (1998). *Détection des observations influentes pour l'Enquête sur les finances des consommateurs (EFC) et l'Enquête sur la dynamique du travail et du revenu (EDTR)*, internal document, Statistics Canada.