ESTIMATING MONTHLY GROSS FLOWS IN LABOUR FORCE PARTICIPATION[1]

Stephen E. Fienberg and Elizabeth A. Stasny[2]

The Canadian Labour Force Survey is a household survey conducted each month for the purpose of producing point-in-time estimates of the number of persons employed, unemployed and not in the labor force. The survey has a rotating panel design in which all individuals in a sampled household location are interviewed each month, for six consecutive months. In the past, little use has been made of this longitudinal structure, although considerable interest has been expressed in the month-to-month gross flows (transitions) amongst the labour force status categories. In this paper we discuss methods being considered by Statistics Canada for the production of gross flow estimates, but from a model-based perspective.

1. INTRODUCTION

The Canadian Labour Force Survey is a monthly household survey used to produce cross-sectional or point-in-time estimates of labour force participation. This survey, however, like the Current Population Survey in the United States and many other large scale sample surveys, is designed using a panel structure so that the subjects are interviewed a number of times before being dropped from the sample. Although the survey is used mainly to obtain cross-sectional estimates, it has long been recognized that information from the repeated interviewing of subjects provides an additional longitudinal data base that could be exploited to give estimates of change over time for a very small additional cost (see, for example Kalachek, 1979, and Fienberg and Tanur, 1983).

---

[1] This research was supported in part by a contract with Statistics Canada. The authors wish to thank Murray Lawes, Larry Swain and Richard Veevers for their help in understanding the Labour Force Survey Methodology and the resulting data, as well as the Editor and a referee for helpful comments and suggestions.

[2] Stephen E. Fienberg and Elizabeth A. Stasny, Carnegie-Mellon University, Pittsburgh, Pennsylvania, PA. 15213.

Some attempts have been made to use the longitudinal data obtained from panel surveys. For example, the longitudinal data from the Current Population Survey has been used since 1948 to produce tables showing gross movements of individuals between labor force categories from one month to the next. Although these tables are produced each month, they have not been published since 1952 because of statistical problems. Smith and Vanski (1979) discuss the production of gross change data using the longitudinal structure of the Current Population Survey.

Recently, Statistics Canada has initiated an investigation of possible uses of the longitudinal data available as a by-product from the Canadian Labour Force Survey. They, too, would like to find a method for producing reliable estimates of gross movements between labour force categories. In this paper, we discuss the methods being considered by Statistics Canada for the production of such gross change data.

In Section 2, we give a brief description of the coverage and design of the Labour Force Survey, and we describe the structure of the resulting data. Then in Section 3 we outline the proposed method for gross flow estimation developed by Statistics Canada, which depends on the used of sample-based weights, adjustment for inflows to and outflows from the population of interest, consistency adjustments, and bias correction for misclassification error. By developing some simple models for the gross flow process, we explore in Section 4 the implications of Statistics Canada's proposed method. Finally, in Section 5 we describe some work on handling non-response in gross flow estimation.

2. DESCRIPTION OF THE LABOUR FORCE SURVEY

2.1 Survey Coverage

Approximately 56,000 households, chosen from the ten provinces of Canada, are included in the Labour Force Survey sample each month. Questionnaires are completed for all civilian, non-institutionalized members of sampled households who are 15 years of age and older. The survey questions primarily relate to the subjects' work related activities during the reference week which is the week prior to the survey week and usually contains the fifteenth day of the month. Responses to the survey questions are used to classify subjects as employed, unemployed, or not in the labour force. For a discussion on classification of labour force status, see Guide to Labour Force Survey Data, Statistics Canada, (1979).

2.2 Survey Design

The labour Force Survey was designed to enable estimation of levels and rates of employment and unemployment for each of the ten provinces separately. Thus, except for the constraint on the total sample size, each province is sampled independently.

Economic regions (ER's), areas of similar economic structure, form the primary strata within provinces. ER's are divided into self-representing units (SRU's) and non-self-representing units (NSRU's). SRU's are large urban centers and NSRU's are generally composed of a small urban center and a rural area. Sampling is carried out separately in SRU's and NSRU's.

SRU's are sampled using a stratified, two-stage sampling design. NSRU's are sampled using a stratified multi-staged sampling scheme. In addition to the SR and NSR areas, some sample units are selected from an apartment frame and a special area frame. The final sampling units for the Labour Force Survey are households. A detailed description of the sampling plan for the survey can be found in Methodology of the Canadian Labour Force Survey 1976, Statistics Canada (1977).

Households selected for the Labour Force Survey are included in the survey for six consecutive months and are then dropped from the sample. For example, households rotated into the survey in January are interviewed for six consecutive months, and then dropped from the sample after the June interview. Each group of households that rotated into and out of the sample together makes up a panel. In any one month, individuals from six different panels are included in the Labour Force Survey sample.

2.3 Sampled-Based Weights

The cross-sectional data, information for a given month from all subjects in the six panels interviewed in that month, is used to produce monthly estimates of labour force participation. The monthly estimates are weighted averages of values for each person in the sample. A weighted average is used because each sampled person is thought of as "representing" a number of people in the population of interest. The weight assigned to an individual's record corresponds to the number of persons in the population that the person in the sample represents.

Let $W_{t,i}$ be the weight assigned to surveyed individual i in the month t. If individual i is classified as outside the population of interest in month t, then $W_{t,i} = 0$. Otherwise, the assigned weights are determined by the probabity of selecting the cluster, the probability of selecting the household within the cluster, nonresponse within the month, rural/urban factors, sub-sampling adjustment for fast-growing areas, and ratio adjustments for province/age/sex factors.

An individual's assigned weight can change from month to month because of replacement of 1/6 of the sample each month, non-response, and, to a lesser extent, because of changes in the size of the population of interest. Thus, for any one individual, i, it may be the case that $W_{t-1,i} = W_{t,i}$.

2.4 Longitudinal Structure and Gross Flow Estimation

Although the main prupose for the Labour Force Survey is to produce point-in-time estimates of labour force participation, the panel structure of the survey design results in a longitudinal data base, with approximately 5/6 of the household locations sampled in any one month being in the sample for the following month. Naturally, fewer than 5/6 of the surveyed individuals or families are the same in two consecutive months due to non-response and moving. However, Statistics Canada is interested in the possibility of using information from those individuals who do respond in two consecutive months to produce estimates of gross flows among labour force categories.

Estimates of gross flows are useful for answering questions such as a) How much of the increase in unemployment is due to persons losing jobs and how much is due to persons formerly not in the labour force starting to look for jobs? or b) How many unemployed persons become discouraged and leave the labour force?

We discuss the problem of estimating gross flows among labour force categories in the next two sections.

3. STATISTICS CANADA'S PROPOSED METHOD FOR GROSS FLOW ESTIMATION

In this section we describe the multi-stage estimation procedure for gross flows developed by Statistics Canada (e.g. see Macredie and Veevers, 1977; Wong, 1983). Our description of the procedure includes various interpretations of the impact of individual stages.

3.1 Data Needed to Estimate Gross Flows

Statistics Canada has proposed estimating gross flows using a 4×4 matrix as shown below:

		Labour Force Status in Month t			
		E	U	N	O
Labour Force Status in Month t-1	E	X_{EE}	X_{EU}	X_{EN}	X_{EO}
	U	X_{UE}	X_{UU}	X_{UN}	X_{UO}
	N	X_{NE}	X_{NU}	X_{NN}	X_{NO}
	O	X_{OE}	X_{OU}	X_{ON}	X_{OO}

where E = employed

U = unemployed

N = not in the labour force

O = outside the population of interest, and

x_{ij} = estimated number with labour force status i in month t-1 and status j in month t.

The final monthly Labour Force Survey files from two consecutive months can be used to obtain the data for estimating the 4×4 matrix of gross flows. In order to use these data to produce gross flow estimates, Statistics Canada must match individual records from the two consecutive monthly files using the unique identification numbers assigned to sampled individuals for the duration of their time in the study.

An individual appearing on the data file for one month may be missing from the file for the other month due to rotation into or out of the sample or because the person moved, was not at home, or refused to respond. The sample weights described in Section 2.4 include an adjustment for non-response within each month. When dealing with gross flows, we also need to consider the month-to-month non-response. Statistics Canada proposes to reweight records for individuals who responded in both months t-1 and t to compensate for this additional non-response.

After the reweighting is completed, Statistics Canada will have a single data file that includes information for all persons who were respondents in two consecutive months. That file will contain geographic and demographic information for each individual as well as the individual's labour force status and assigned weights for both month t-1 and month t.

3.2 Differences in Weights

As we noted in Section 2.3, an individual's sample-based weight can change from month to month because of the rotation replacement structure, non-response, and changes in the size of the population of interest. Even when the adjustment factor for non-response is computed on the basis of month-to-month data, for any individual i, it may still be the case that $W_{t-1,i} \neq W_{t,i}$. Some method is needed for handling this difference in weights if data are to be used for estimating gross flows.

Statistics Canada proposes to resolve this dilemma by assuming that differences in the two weights occur only as a result of inflows to and outflows from the population of interest. Thus, differences in weights are added to the appropriate cell in either the last row or last column of the gross flow matrix. This procedure depends heavily on the interpretation of the weights suggested in Section 2.3, namely that sample individual i represents $W_{t,i}$ persons in the population in month t.

As an illustration of the procedure, suppose an individual is classified as employed in both months t-1 and t but $W_{t-1,i} = 300$ and $W_{t,i} = 305$. The minimum weight, 300, is added to the EE cell of the gross flow table. The difference, $W_{t,i} - W_{t-1,i}$, of 5 is added to the OE cell since those 5 would be thought of as having been outside the population of interest during the month t-1 and then having moved in the population as employed persons for month t.

If, on the other hand, the individual is employed in both months but $W_{t-1,i}$ = 305 and $W_{t,i}$ = 300, then 300 is again added to the EE cell but the excess weight of 5 is added to the EO cell. Here, the difference between weights represents 5 persons who were employed in month t-1 and then moved outside the population of interest in month t.

An individual who is classified as outside the population of interest in month t-1 and is then, say, employed in month t will have $W_{t-1,i}$ = 0. If $W_{t,i}$ =300, then 300 is added to the OE cell. Individuals classified as outside the population of interest in month t are treated similarly with $W_{t-1,i}$, being added to the appropriate cell in the last column of the gross flow matrix.

Because persons outside the population of interest are assigned a weight of zero, a person who is classified as such in both months t-1 and t would have $W_{t-1,i}$ = 0 and $W_{t,i}$ = 0. Therefore, X_{OO}, the entry in the OO cell of the gross flow matrix, must always be zero.

3.3 Adjustment of the Inflow and Outflow Cells

Adding differences in weights to the inflow and outflow cells of the gross flow matrix provides a method for handling the changes in sample-based weights from one month to the next and gives estimates of inflows to and outflows from the population of interest. Independent estimates of inflows and outflows, available from Census data, suggest that this method overestimates the actual amount of movement into and out of the population of interest. Thus, Statistics Canada plans to adjust the X_{OE}, X_{OU}, X_{ON}, X_{EO}, X_{UO}, and X_{NO} entries in the gross flow matrix. These cells will be proportionally adjusted so that total inflows and outflows shown in the gross flow matrix equal the Census estimates of inflows and outflows respectively.

Let I be the independent census estimate of inflows to the population of interest and F be the census estimate of outflows from the population. Call the sum of estimated inflows $X_{0+} = X_{0E} + X_{0U} + X_{0N}$ and the sum of estimated outflows $X_{+0} = X_{E0} + X_{U0} + X_{N0}$. The proportionally adjusted inflows are:

$$Y_{0j} = X_{0j} I / X_{0+} \qquad \text{for } j = E, U, N. \qquad (1)$$

The proportionally adjusted outflows are

$$Y_{j0} = X_{j0} F / X_{+0} \qquad \text{for } j = E, U, N. \qquad (2)$$

3.4 Consistency of Gross Flow Estimates With Monthly Totals

Statistics Canada would like their gross flow estimates to be consistent with the published monthly estimates of labour force participation totals. Thus, the row totals for the gross flow matrix must be the month t-1 estimates of labour force participation and the column totals must be the month t cross-sectional estimates. The marginal totals of the gross flow matrix constructed as described above are not consistent with the monthly labour force totals.

Statistics Canada plans to use the method of iterative proportional scaling, originally proposed by Deming and Stephan (1940), and described in detail by Bishop, Fienberg, and Holland (1975), to adjust the gross flow matrix to agree with the monthly labour force totals. When used to adjust the gross flow matrix, iterative proportional scaling alternatively 1) constrains the rows of the matrix to sum to the month t-1 estimates and then 2) constrains the columns to sum to the month t estimates. Steps 1) and 2) are repeated until the entries in the matrix do not change from one step to the next.

Testing at Statistics Canada has shown that cell changes resulting from the application of iterative proportional scaling were both absolutely and relatively small and fell roughly within the bounds of sampling variability associated with the cells. This suggests that the consistency adjustment does not seriously distort the gross flow estimates.

3.5 Bias Correction for Misclassification Error

Statistics Canada proposed method for estimating gross flows also includes a step correcting for misclassification bias. This is the bias that results from the incorrect assignment of an individual's labour force status. A technique developed by Fred Wong (1983) at Statistics Canada uses reinterview data to correct for the misclassification bias.

4. IMPLICATIONS OF STATISTICS CANADA'S PROPOSED METHOD

4.1 Modeling Gross Flows

Each step of Statistics Canada's proposed method described in the previous section is a logical attempt to correct problems that arise concerning the production of good estimates of gross flows. It is not clear, however, what effect the various adjustments have on the final estimated gross flow matrix. In order to better understand Statistics Canada's proposal to treat differences in weights as being due to inflows to and outflows from the population of interest, in this section we develop a model for the gross flow process. Our discussion centers on the quantities in the inflow and outflow cells of the estimated gross flow matrix since the problems with Statistics Canada's proposed method seem to occur in those cells. Because the design of the Canadian Labour Force Survey is quite complex, we begin with a set of simplifying assumptions. In the following we assume that

1. a single stage stratified sample is chosen,
2. there is no response error, and
3. non-response occurs only because random individuals move between strata or because of rotation into or out of the sample.

4.2 Allocation of Net Population Changes to Inflow and Outflow Cells

Suppose that the population of interest is divided into S strata indexed by $s = 1,2, \dots S$. Let

N_k^s = population size in strata s in month k.

Each month, a simple random sample is chosen from each stratum for the survey and sampled individuals are interviewed for six consecutive months before being dropped from the survey. Our goal is to estimate gross flows from month t-1 to month t.

For the purpose of estimating gross flows, only individuals who are interviewed in both month t-1 and t will be used. This excludes individuals who rotate into or out of the sample and persons who move between strata. Let

$r_{t-1,t}^s$ = number of sampled individuals from stratum s who were interviewed in both months t-1 and t.

Each of the $r_{t-1,t}^s$ respondents in stratum s is assigned the following weights in months t-1 and t respectively for the purpose of gross flow estimation:

$$W_{t-1}^s = N_{t-1}^s \;/\; r_{t-1,t}^s \text{ and } W_t^s = N_t^s \;/\; r_{t-1,t}^s . \qquad (3)$$

As long as movements between strata and selection of panels are "random processes," these weights represent the inverse of the probability that an individual within a stratum is interviewed in both months t-1 and t. Since all individuals within a stratum have the same weight in any given month, aggregates for each stratum may be used. Therefore, we let

n_{ij}^s = number of sampled individuals from stratum s classified as having labour force status i in month t-1 and status j in month t for i,j = E,U,N,O.

The methodology proposed by Statistics Canada requires that the minimum of the months t-1 and t weights for each individual be added to the appropriate cell in the gross flow matrix. The difference is added to the appropriate inflow cell if the month t weight is greater than the weight in month t-1 and to the appropriate outflow cell otherwise. Thus, for example, the stratum s entry in the EE (employed to employed) cell of the gross flow matrix is:

$$\min(W^s_{t-1}, W^s_t) n^s_{EE} = \min \left[(N^s_{t-1} / r^s_{t-1,t}), (N^s_t / r^s_{t-1,t})\right] n^s_{EE}$$

$$= \min (N^s_{t-1}, N^s_t)\, n^s_{EE} / r^s_{t-1,t}$$

$$= \min (N^s_{t-1}, N^s_t)\, f^s_{EE} \qquad (4)$$

where f^s_{EE} = fraction of all individuals from stratum s, interviewed in both months t-1 and t, who where employed in both months.

The contribution from stratum s to the OE cell for the matrix from individuals employed in month t is:

$$\max(0, W^s_t - W^s_{t-1}) n^s_{EE} = \max\left[0, (N^s_t - N^s_{t-1} / r^s_{t-1,t})\right] n^s_{EE}$$

$$= \max(0, N^s_t - N^s_{t-1})\, n^s_{EE} / r^s_{t-1,t}. \qquad (5)$$

Differences from individuals falling in the UE and NE cells will also contribute to the OE cell. Thus, the total contribution to the OE cell from stratum s is:

$$\max(0, N^s_t - N^s_{t-1}) \left\{(n^s_{EE} / r^s_{t-1,t}) + (n^s_{UE} / r^s_{t-1,t}) + (n^s_{NE} / r^s_{t-1,t})\right\}$$

$$= \max(0, N^s_t - N^s_{t-1})\, n^s_{+E} / r^s_{t-1,t}$$

$$= \max(0, N^s_t - N^s_{t-1})\, f^s_{+E} \qquad (6)$$

where f^s_{+E} = fraction of all individuals from stratum s, interviewed in both months t-1 and t, who were employed in month t.

We obtain total for all cells in the gross flow matrix in a similar manner. The resulting gross flow matrix is as follows:

Gross Flow Matrix - Month t-1 to Month t

Month t-1 \ Month t	E	U	N	O	
E	$\sum_{s=1}^{S} \min(N^s_{t-1}, N^s_t) f^s_{EE}$	$\sum_{s=1}^{S} \min(N^s_{t-1}, N^s_t) f^s_{EU}$	$\sum_{s=1}^{S} \min(N^s_{t-1}, N^s_t) f^s_{EN}$	$\sum_{s=1}^{S} \max(0, N^s_{t-1} - N^s_t) f^s_{E+}$	$\sum_{s=1}^{S} N^s_{t-1} f^s_{E+}$
U	$\sum_{s=1}^{S} \min(N^s_{t-1}, N^s_t) f^s_{UE}$	$\sum_{s=1}^{S} \min(N^s_{t-1}, N^s_t) f^s_{UU}$	$\sum_{s=1}^{S} \min(N^s_{t-1}, N^s_t) f^s_{UN}$	$\sum_{s=1}^{S} \max(0, N^s_{t-1} - N^s_t) f^s_{U+}$	$\sum_{s=1}^{S} N^s_{t-1} f^s_{U+}$
N	$\sum_{s=1}^{S} \min(N^s_{t-1}, N^s_t) f^s_{NE}$	$\sum_{s=1}^{S} \min(N^s_{t-1}, N^s_t) f^s_{NU}$	$\sum_{s=1}^{S} \min(N^s_{t-1}, N^s_t) f_{NN}$	$\sum_{s=1}^{S} \max(0, N^s_{t-1} - N^s_t) f^s_{N+}$	$\sum_{s=1}^{S} N^s_{t-1} f^s_{N+}$
O	$\sum_{s=1}^{S} \max(0, N^s_t - N^s_{t-1}) f^s_{+E}$	$\sum_{s=1}^{S} \max(0, N^s_t - N^s_{t-1}) f^s_{+U}$	$\sum_{s=1}^{S} \max(0, N^s_t - N^s_{t-1}) f^s_{+N}$	0	
	$\sum_{s=1}^{S} N^s_t f^s_{+E}$	$\sum_{s=1}^{S} N^s_t f^s_{+U}$	$\sum_{s=1}^{S} N^s_t f^s_{+N}$		

Notice that each term in the summations for the nine in-population cells (the cells showing gross flows between employed, unemployed, and not in the labour force) is the product of the net size of the strata and the observed fraction of subjects who had the various labour force classifications in months t-1 and t. The out-of-population to employed cell of the gross flow matrix contains a sum of terms from each strata that grew from month t-1 to month t. Each term is the product of the net increase in size for the strata and the fraction of the subjects in the strata who reported being employed in month t.The out-of-population to unemployed and not in the population of interest cells contain the sums of similar terms except that the net increase in size for each strata is multiplied by the fraction of subjects in the strata who were unemployed or not in the labour force in month t respectively. In other words, the net increase in size for each strata is proportionally allocated to the three inflow cells of the gross flow matrix based on the observed fractions of employed, unemployed, and not in the labour force in month t. Similarly, the net decrease in size for each strata that shrank between months t-1 and t is proportionally allocated to the outflow cells of the matrix based on the observed fractions of employed, unemployed, and not in the labour force in month t-1.

In this model we have assumed that the only way for counts to appear in the inflow and outflow cells is as a result of differences in weights. In practice, a small number of individuals who move in and out of the population-of-interest show up in the sample and their assigned weights are added to the appropriate inflow and outflow cells. The effect of such individuals on the estimates is very small.

The fractions f^s_{+E}, f^s_{+U}, f^s_{+N}, f^s_{E+}, f^s_{U+}, and f^s_{N+} are estimated using individuals who appear in the sample in both months. Almost all individuals classified, for example, as OE could not be respondents in both months because they are not sampled by design or because they are movers. Thus, these people who could not have been respondents in both months are represented by people who did respond in both months. To the extent that these groups differ, the proportional allocation of net increases and decreases in strata size may result in biased estimates in the inflow and outflow of the gross flow matrix.

4.3 Effects of Movements Between Strata

The weights used for the purpose of gross flow estimation, as shown in expression (3), are determined by the number of respondents in both months t-1 and t, a quantity that remains constant for the two months, and by the stratum population. The population of a stratum changes if a) individuals enter from outside the population of interest, such as when persons reach their 15th birthday or leave the full-time military, b) individuals move outside the population of interest, as when persons enter the military or an institution, or c) individuals in the population of interest move between strata. This subsection describes the effects of such changes in population size on the quantities in the gross flow matrix.

As in the preceding subsection, we suppose that the population of interest is divided into S strata. Again, individuals are sampled at random from each stratum every month, interviewed for six consecutive months, and then dropped from the sample. Let $r^s_{t-1,t}$ be, as before the number of individuals from stratum s who are interviewed in both months t-1 and t.

Next we suppose that there are N^s_{t-1} individuals in stratum s in month t-1. Let movements into and out of strata between months t-1 and t be denoted by

$m_{u,v}$ = number of individuals who move from u to v, u ≠ v, between interviews for months t-1 and t where u and v may take on the values.

s = stratum s for s = 1,2, ... S and

0 = outside the population of interest.

Using this notation, the population in stratum s in month t is

$$N^s_t = N^s_{t-1} + \Sigma_{u \neq s} (m_{u,s} - m_{s,u}). \qquad (7)$$

The weights assigned to individuals in stratum s in months t-1 and t respectively are

$$W^s_{t-1} = N^s_{t-1} / r^s_{t-1,t} \text{ and } W^s_t = N^s_{t-1} + \Sigma_{u \neq s} (m_{u,s} - m_{s,u}) / r^s_{t-1,t}. \qquad (8)$$

Since our focus in this section is on movement into and out of the population of interest, it is not necessary for us to divide those in the population of interest into employed, unemployed, and not in the labour force. Thus, the gross flow matrix used here is a 2×2 matrix formed by collapsing the first three rows and columns of the 4×4 gross flow matrix used in the preceding sub-section.

The entry for stratum s in the in-population to in-population cell is

$$\min(W^s_{t-1}, W^s_t) r^s_{t-1,t} = \min\{N^s_{t-1}/\ r^s_{t-1,t},\ [N^s_{t-1} + \Sigma_{u\neq s}(m_{u,s} - m_{s,u})]/r^s_{t-1,t}\} r^s_{t-1,t}$$

$$= \min[N^s_{t-1},\ N^s_{t-1} + \ \Sigma_{u\neq s}\ (m_{u,s} - m_{s,u})]$$

$$= N^s_{t-1} + \min[0,\ \Sigma_{u\neq s}\ (m_{u,s} - m_{s,u})]. \qquad (9)$$

The entry for stratum s in the out-of-population to in-population, or inflow, cell is

$$\max(0, W^s_t - W^s_{t-1}) r^s_{t-1,t} = \max[0, \Sigma_{u\neq s}\ (m_{u,s} - m_{s,u})/r^s_{t-1,t}]\ r^s_{t-1,t}$$

$$= \max[0, \Sigma_{u\neq s}\ (m_{u,s} - m_{s,u})]. \qquad (10)$$

The entry for the in-population to out-of-population, or outflow, cell is found similarly. Thus, the 2×2 gross flow matrix is as follows:

		Month t: In-Population	Month t: Out-of-Population	
Month t-1	In population	$\sum_{s=1}^{S} \{N_{t-1}^{s} + \min[0, \Sigma_{u\neq s}(m_{u,s} - m_{s,u})]\}$	$\sum_{s=1}^{S} \max[0, \Sigma_{u\neq s}(m_{u,s} - m_{s,u})]$	$\sum_{s=1}^{S} N_{t-1}^{s}$
	Out-of-population	$\sum_{s=1}^{S} \max[0, \Sigma_{u\neq s}(m_{u,s} - m_{s,u})]$	0	

$$\sum_{s=1}^{S} \{N_{t-1}^{s} + \Sigma_{u\neq s}(m_{u,s} - m_{s,u})\}$$

Let us consider the quantity in the inflow cell of this gross flow matrix. This cell should contain the net increase in population from outside the population of interest, $m_{0,s} - m_{s,0}$, for each stratum that gained members from outside the population. What the cell does contain is $\Sigma_{u\neq s}(m_{u,s} - m_{s,u})$ for each stratum s that grew as a result of movements between strata and from outside the population of interest. The summation, $\Sigma_{u\neq s}(m_{u,s} - m_{s,u})$, does include the quantity $m_{0,s} - m_{s,0}$ but it may also contain other terms.

For example, suppose the population is made up of three strata called A, B, and C. If strata A and B grew from month t-1 to month t and stratum C lost members, then the inflow cell contains

$$\Sigma_{u\neq A}(m_{u,A} - m_{A,u}) + \Sigma_{u\neq B}(m_{u,B} - m_{B,u}) = m_{0,A} - m_{A,0} + m_{B,A} - m_{A,B} + m_{C,A} - m_{A,C}$$

$$+ m_{0,B} - m_{B,0} + m_{A,B} - m_{B,A} + m_{C,B} - m_{B,C}$$

$$= m_{0,A} - m_{A,0} + m_{0,B} - m_{B,0} + m_{C,A} - m_{A,C} + m_{C,B} - m_{B,C} \quad (11)$$

Note that the movements between strata A and B cancel out but the terms showing the movement between strata A and C and strata B and C remain in the summation.

In general, the inflow cell contain extra terms of the form $m_{v,u}-m_{u,v}$ for each stratum v that loses population while stratum u gains population. Similarly, the outflow cell contains extra terms of the form $m_{x,y}-m_{y,x}$ for each stratum y that grows while stratum x loses population.

In the inflow cell, the quantity $\Sigma_{u \neq s}(m_{u,s}-m_{s,u})$ for each strata s that gains population from month t-1 to t will be positive, although each individual term in the summation need not be positive. If

$$\Sigma_{u \neq s} (m_{u,s} - m_{s,u}) > m_{0,s} - m_{s,0} \tag{12}$$

then the contribution to stratum s is more than the inflow to stratum s from outside the population of interest. This excess comes from terms of the form $m_{u,v} - m_{v,u}$ as described above. That is, the overestimate is due to movements between strata within the population. A similar result holds for the in-population to out-of-population cell of the matrix.

Statistics Canada staff report that the method they proposed for handling differences in weights from month to month does appear to give overestimates in the inflow and outflow cells of the gross flow matrix. Although they are based on simplifying assumptions, the results here give a possible explanation for the overestimation, i.e. the overestimation may be due to movements within the population of interest.

Finally we note that, in the 2×2 gross flow matrix shown above, the in-population to in-population cell must contain an underestimate equal to the overestimate in the outflow cell. Whatever the amount of underestimation, it is spread over the nine in-population to in-population cells in the 4×4 gross flow matrix. Moreover, the size of the overestimation is small in comparison to the total size of the nine in-population cells.

4.4 Comments on the Proposed Gross Flow Estimation Method

The results described in the preceding two subsections illustrate problems with the proposed method of handling month-to-month differences in weights for the purpose of gross flow estimation. These results do not come as a surprise to Statistics Canada. Because of their experience with Labour Force Survey methods and data, they realized that the movements in individuals within the population might explain some of the overestimation in the inflow and outlfow cells of the gross flow matrix. The results obtained by modelling the process reinforce their beliefs and make it clear just how the movements of individuals effects the estimates. In addition, the modelling brought to light a problem about which Statistics Canada had not been aware: the compensating underestimation spread over the nine in-population to in-population cells of the gross flow matrix.

In section 4.2, we saw that the net increases in strata are allocated to the inflow cells while the net decreases are allocated to the outflow cells according to the fractions of observed individuals classified as employed, unemployed, and not in the labour force in month t and month t-1, respectively. The estimation of inflows and outflows in this manner is valid only if individuals who move in and out of the population of interest are a random sample of individuals and, hence, "the same" as individuals who remain within the population of interest. Sampled individuals who are classified as outside the population of interest appear in the sample by accident rather than by design; the Labour Force Survey is not designed to estimate numbers of persons outside the population of interest. If we want to obtain reasonable estimates for the inflow and outflow cells of the matrix, it may be necessary to include individuals outside the population of interest in the Labour Force Survey sample or to use a special, supplementary sample.

In section 4.3, we saw that the overestimates in the inflow and outflow cells could be a result of movements of individuals between a strata whose population grew and a strata whose population shrank. The fact that it was movements between strata that caused the problem is a result of the simplifying assumptions which we made. We assumed that the final sample was randomly chosen from within each strata. Hence, the weights assigned to individuals sampled from a single strata were equal. If, instead, we assumed that the strata had been divided into clusters and random samples of individuals had been chosen from within the clusters, then all individuals sampled from a single cluster would have been assigned the same weight and the overestimate would come as a result of movements between clusters.

To correct for the overestimate, and corresponding underestimate, directly in the case where final samples are chosen at random from within strata, we would need estimates of the number of movers between each pair of strata where one strata grew and the other strata lost population. If the final samples are chosen at random from within clusters, similar estimates would be required for each pair of clusters. This is a considerable amount of information. A further complication is that, in practice, the ratio adjustments applied to the weights make it possible for individuals within one household to have unequal weights.

As was suggested earlier, if individuals outside the population of interest were included in the sample, we could obtain estimates of movement into and out of the population of interest directly. One other possibility that should be considered is to discard the monthly weights for the purpose of gross flow estimation and derive a longitudinal weight for each individual in the Labour Force Survey sample in either of the two months.

As statisticians, we are quite comfortable with estimates of gross flows that do not have the published monthly labour force participation totals as marginal totals; however, we realize the problems that might arise if gross flow estimates, that were not consistent with the monthly totals, were published. Nevertheless, it should not be assumed automatically that the monthly estimates are correct and that the problem lies solely in the gross flow estimates.

As we noted in section 3.5, the gross flow matrix is adjusted to correct for misclassification errors. The monthly estimates, however, are not corrected for misclassification bias. Thus, when iterative proportional fitting is used to adjust the gross flow matrix to agree with the monthly totals, the matrix is being altered to be consistent with biased values. We feel that it would be more appropriate to address the problem of misclassification of labour force status in the monthly data where it occurs rather than just in the estimates of gross flows.

5. NON RESPONSE AND GROW FLOW ESTIMATION

Statistics Canada's proposed method for gross flow estimation compensates for non-response by adjusting the sample-based weights of respondents. This method of handling non-response is appropriate if the missing data are missing at random (e.g., see Rubin, 1976). In order to explore the assumption of random non-response, we used a longitudinal file for a single panel to produce the data in Table 1. This table shows the unweighted percentages of individuals reporting being employed or unemployed in zero to six months according to the number of months in which they responded to the survey.

Consider the probabilities underlying the observed percentages shown in part (a) of Table 1. Let

π_i = probability that an individual is employed in i out of 6 months for i = 0,1,... 6.

Under the assumption that non-response occurs at random, the probabilities corresponding to the first column of that table can be written as

P(observing 0 months employed out of 6-k months responding)

$$= \sum_{j=1}^{k} \pi_j \Big/ \binom{6}{j}, \quad \text{for } k = 0,1,\ldots 5. \qquad (13)$$

Notice that these probabilities increase from the first row of the column to the last row.

In a similar manner, it can be shown that, if data are missing at random, then the underlying probabilities must increase from the top to bottom of each column in both tables. The first column of each table deviates from this pattern quite noticeably. In both cases, the observed percentages decrease through the first four rows of the table and then increase in the last two rows. It does not seem likely that sampling variability alone could be responsible for such a pattern in both tables. Thus, it appears as if there is some evidence that non-response does not occur at random.

Of course, the above analysis is based on just a single panel of Labour Force Survey data. However, in a larger study using data from 1980 and 1981, Paul and Lawes (1982) also found evidence of a relationship between employment status and non-response. Therefore, there is a need to consider methods for gross flow estimation that do not require the assumption that non-response occurs at random.

Statistics Canada's proposed method for gross flow estimation only makes use of the information from individuals who responded in both of the months. There is also information available from those individuals who responded in just one of the two months. Stasny (1983) presents a method for month-to-month gross flow estimation that makes use of the information available from individuals who are respondents in only one of the two months and that can be used when non-response is related to time or employment status. For this method, we take the observed gross flow data to be the end result of a two-stage process. In the first stage of the process, which we do not get to observe, individuals are allocated to the sixteen cells of the gross flow matrix according to a single multinomial sampling scheme. Then, in the second stage, each individual may lose either the month t-1 or month t labour force classification with some probability. The probability of losing a month's classification can be modeled to depend on the month, or labour force status, or both. Maximum likelihood estimates for the parameters of the multinomial distribution of the first stage and the probabilities of losing a month's classification are obtained using iterative methods.

When these models were fit to Labour Force Survey data from a single panel, Stasny (1983) found that the model where the probability of losing a month's classification depends on labour force status provides a reasonable fit to the data for all gross flow matrices with the exception of the months 1-2 matrix. For the data from month 1 to month 2, the probability of losing a month's classification appears to depend on the month. This may be due to the fact that there is higher non-response in the first month a panel is in the survey. We believe that it would be worthwhile to fit this type of model to additional data from the Labour Force Survey to see if similar results are obtained over other panels.

Clearly, the problem of obtaining good estimates of gross flows from the Labour Force Survey is not a simple one. The survey is designed to give data for the production of monthly estimates of labour force participation, not estimates of gross flows. A survey designed specifically for the purpose of estimating gross flows among labour force categories would certainly be different from the current Labour Force Survey. Thus, the longitudinal data from the survey is not ideal for gross flow estimation. The data, however, are available and, if they can be used to give reasonable estimates of gross flows, then additional, useful information is produced for a relatively small cost.

REFERENCES

[1] Bishop, Y.M.M., Fienberg, S.E., and Holland, P.W. Discrete Multivariate Analysis: Theory and Practice. The MIT Press, 1975.

[2] Deming, W.E. and Stephan, F.F. On A Least Squares Adjustment Of a Sampled Frequency Table When the Expected Marginal Totals Are Known. Annals of Mathematical Statistics, 1940, 11, 427-444.

[3] Fienberg, S.E. amd Tanur, J.M. The Design and Analysis of Longitudinal Surveys: Controversies and Issues of Cost and Continuity. Technical Report 289, Department of Statistics, Carnegie-Mellon University, May 1983.

[4] Kalachek, E. Longitudinal Surveys and Labor Market Analysis. Data Collection, Processing and Presentations: National and Local (Counting the Labor Force, Appendix, Volume II), U.S. Government Printing Office, 1979, 160-189.

[5] Macredie, I.D. and Veevers, R. Discussion on Smith, R.E. and Vanski, J.E. Gross Change Data: The Neglected Data Base. Data Collection, Processing and Presentation: National and Local (Counting the Labor Force, Appendix, Volume II), U.S. Government Printing Office, 1979, 153-158.

[6] Paul, E.C. and Lawes, M. Characteristics of Respondent and Non-Respondent Households in the Canadian Labour Force Survey. Survey Methodology, 1982, 8, 48-85.

[7] Rubin, D.B. Inference and Missing Data. Biometrika, 1976, 63, 581-592.

[8] Smith, R.E. and Vanski, J.E. Gross Change Data: The Neglected Data Base. Data Collection, Processing and Presentation: National and Local (Counting the Labor Force, Appendix, Volume II), U.S. Government Printing Office, 1979, 132-150.

[9] Stasny, E.A. Estimating Gross Flows in Labor Force Participation Using Information From Individuals With Incomplete Classifications. Technical Report 272, Department of Statistics, Carnegie-Mellon University, Jan. 1983.

[10] Statistics Canada, Guide to Labour Force Survey Data. Catalogue 71-528 Occasional, July 1979.

[11] Statistics Canada, Methodology of the Canadian Labour Force Survey 1976. Catalogue 71-526 Occasional, Oct. 1977.

[12] Wong, F.A. Technique To Correct The Response Bias In the 4×4 Labour Force Gross Flow Matrix, Technical Reprot, Statistics Canada, 1983.

Table 1

a)

Months of Data Present	Months Employed 0	1	2	3	4	5	6
6	52.23	3.22	1.93	2.21	2.45	3.55	34.41
5	51.48	3.04	2.17	3.36	4.22	35.73	
4	49.26	4.15	3.16	4.83	38.60		
3	46.31	6.18	5.62	41.89			
2	51.40	8.32	40.28				
1	52.87	47.13					

b)

Months of Data Present	Months Unemployed 0	1	2	3	4	5	6
6	92.43	3.63	1.65	0.93	0.57	0.43	0.35
5	91.23	4.46	2.08	1.13	0.72	0.38	
4	89.28	5.76	2.17	1.43	1.36		
3	89.00	6.42	2.81	1.77			
2	91.33	6.01	2.66				
1	91.43	8.57					