COMPARISON OF SOME RATIO TYPE ESTIMATORS FOR LARGE SCALE HOUSEHOLD SURVEYS[1]

M. Lawes and M.P. Singh[2]

In this paper three types of ratio estimators, namely combined, post-stratified and a generalized ratio estimator developed earlier by Singh (1969) and Naga Reddy (1974), are considered. Based on an empirical evaluation, their efficiencies are compared for two large scale household surveys, namely the Canadian Labour Force Survey and the Survey of Consumer Finances.

1. INTRODUCTION

Stratified multistage sample design is usually adopted for conducting large scale household surveys due to its operational convenience. Data available on auxiliary variables are utilized at various stages of sampling operations such as formation of strata and sampling units, determination of overall sample size and its allocation to different strata and stages of sampling, sample selection using unequal probabilities, etc. Auxiliary data needed for performing these sampling operations are usually at lower levels for example for each sampling unit or even for each element (i.e. blocks or enumeration areas) that comprise sampling units. Quite often, additional data are available from independent sources but at higher levels of aggregations (i.e. by census divisions or by provinces) and by different classifications

---

1 This paper was presented at the meeting of the Statistical Society of Canada, Montreal, May 27-30, 1980.

2 M. Lawes and M.P. Singh, Census and Household Survey Methods Division, Statistics Canada.

(i.e. population by age, sex, occupation, household, family type, marital status, etc.). Such data may be used efficiently at the estimation stage for constructing ratio or regression estimators.

In this paper, three types of ratio estimators are considered and their efficiencies are compared for two large scale household surveys, namely, the Canadian Labour Force and the Survey of Consumer Finances. The estimators examined are: a combined ratio estimator using data at province level, post-stratified ratio estimator using data by different sub-classes at province level and a generalized ratio estimator developed by Singh [1969] and Reddy [1974]. These estimators, along with simple survey estimator, are discussed in section 3 after giving a brief description of the two surveys in the following section. Using Keyfitz [1957] method for variance estimation, the expressions for the variances are given in section 4 under the usual approximation of ratio estimation. The efficiency comparison and analysis of results are then presented in the last section.

2. BRIEF DESCRIPTION OF THE SURVEYS

The estimation procedures have been applied to and evaluated for two surveys conducted by Statistics Canada, namely, the Canadian Labour Force Survey and the Survey of Consumer Finances. A detailed description of the Labour Force Survey is presented in Methodology of the Canadian Labour Force Survey (1977). Although the samples are selected from a common sampling frame, the type of data collected varies between the two surveys. For the Labour Force Survey described in sub-section 2.1, the majority of the data items are qualitative in nature dealing with the labour force activities of the repsondent during reference week whereas for the Survey of Consumer Finances discussed in sub-section 2.2, the data items are primarily quantitative in nature consisting of income amounts by source of income for the previous calendar year.

2.1 The Labour Force Survey

The Labour Force Survey is conducted on a monthly basis to collect information on the work activities and employment status of Canadians. The survey covers the civilian non-institutional population in the 10 provinces of Canada excluding residents of Indian Reservations and members of the Canadian Armed Forces. Each month a sample of about 62,000 dwellings is selected based on a multistage stratified probability sampling procedure. Information on the Labour Force activities of all eligible household members 15 years of age or over is collected by interviewers by either a personal or telephone interview. Within geographically contiguous economic regions the sample is selected independently within the following three types of areas:

a) Self-representing areas are comprised of the larger cities. Each city is divided into sub-units whose size varies from 1,000 to 12,000 households depending on the sampling ratio for the province. Sub-units so formed are subdivided into clusters which are usually city blocks. The clusters are randomly grouped into multiples of six groups, to facilitate the implementation of six-month household rotation scheme.

   One cluster per group is selected with probability proportional to size and a systematic sample of dwellings is selected from within selected clusters.

b) Non self-representing areas consist of rural and small urban areas. These areas are initially divided into strata containing on the average about 15 primary sampling units (PSUs). Two or more PSUs per stratum are selected with probability proportional to size. At the subsequent stages, subsampling is carried out independently within urban and rural parts of the selected PSUs.

c) Special areas are composed of areas or establishments which possess characteristics differing from the general population and which may require special interviewing techniques. Included in the special areas are hospitals, schools, hotels, military establishments and remote areas. Special areas make up about 1% of the sample frame.

2.2 The Survey of Consumer Finances

The Survey of Consumer Finances is conducted on an annual basis in the ten provinces of Canada with the exclusion of inmates of institutions and residents of Indian Reservations. There are two variants to the survey. In even numbered years the survey is conducted as a supplement to the April Labour Force Survey. A subsample of about 41,000 dwellings is selected from the Labour Force Survey sample for that month. The subsampling is generally carried out by selecting all sampled households in a subset of the selected clusters. The subject matter content for the supplement is restricted to basic questions on income and work experience during the previous calendar year. In odd numbered years an independent sample of about 17,000 dwellings is selected from the Labour Force Survey sampling frame following the basic sampling procedure described for the Labour Force Survey. In addition to information collected in the even numbered years, data on income related topics are collected in odd numbered years.

3. ESTIMATION PROCEDURE

Although the estimation procedures used for the two surveys differ in terms of minor details, the basic approaches are similar. Survey weights attached to respondent records are the product of six factors, the product of the first five being termed the subweight as detailed below:

a) the basic weight - In a probability sample, the sample design itself determines the weights which may be used to produce unbiased estimates. For both these surveys, using self-weighting design, the weights are the same within each type of area in a province and are equal to the inverse of the sampling ratio.

b) the rural-urban factor - This factor is relevant for non self representing areas, and adjusts sample distributions between rural-urban areas of selected PSU's within provinces to agree with distributions obtainable from census data.

c) balancing factor for non-response - The factors are calculated by the type of area (rural-urban) within primary sampling units and within subunits for self-representing areas.

d) the cluster weight - Within sampled clusters experiencing significant growth between the time of design and interviewing, subsampling of the cluster is carried out to avoid disruptions in the field operations. If subsampling has been carried out within a cluster, the cluster weight which is the inverse of the cluster subsampling rate, is applied to selected households in the cluster.

e) the sample size stabilization weight - To prevent uncontrolled growth in the sample size of the Labour Force Survey, a sample stabilization procedure is introduced into the survey operations under which a number of dwellings that are in excess of the predetermined level are dropped and a stabilization weight is used to compensate for the reduction in the total sample size. For the Survey of Consumer Finances, the sample stabilization capabilities have been used to effect a subsampling of households within selected clusters. The product of the above five factors is termed the subweight.

Mathematically, the subweight may be defined as:

$$Wphijc = Wph\ Fpj\ Bphij\ Cphic\ Spj$$

where

p	denotes province
h	denotes stratum
i	denotes primary sampling unit
j	denotes type of area (SRU, special, urban or rural)
c	denotes cluster
Wph	is the basic weight
Fpj	is the rural-urban factor
Bphij	is the non-response balancing factor
Cphic	is the cluster weight
Spj	is the sample size stabilization weight.

Simple survey estimates of aggregate totals can then be expressed as

$$\hat{X}_p^{(1)} = \sum_{hijc}\sum\sum\sum \; \sum_{k\varepsilon(hijc)} Wphijc \quad Xphijck \qquad (3.1)$$

where

$\sum_{k\varepsilon(hijc)}$ denotes the summation over all records corresponding to respondent sample units in the area identified by (hijc).

and Xphijck denotes the value for record k in (phijc).
For qualitative characteristics, Xphijk is an indicator variable with a value 1 if the record possesses the characteristic under question and zero otherwise. For quantitative characteristics, Xphijk is the value as reported on the record.

Simple survey estimates can be derived on the basis of these subweights. For qualitative characteristics, simple survey estimates of the total number of units possessing a characteristic are derived as the sum of the subweights on all records possessing the characteristic. For quantitative characteristics, the sum of the product of the subweights and reported values yields a simple survey estimate of the characteristic total.

As is well known, the efficiency of the final estimator compared to the simple survey estimator can be increased by utilizing additional information on a related characteristic in the form of a ratio estimation. We define below the three types of ratio estimators that are compared in this article. All of these use data on inter-censal population estimates available at the province level.

Let $\hat{P}_p$ and $\hat{X}_p$ denote simple survey estimates of the population and characteristic totals respectively at the provincial level. Note that $\hat{X}_p$ is the same as $\hat{X}_p^{(1)}$ defined in (3.1), the superscript has been dropped for the sake of convenience in defining the preceeding estimators, and $\hat{P}_p$ is determined by taking the simple sum of weights in (3.1), i.e. assuming x's to be 1. The auxiliary information, P_p the total population count for the province, is obtained from an external source independent of the survey data. The first type of ratio estimator considered is a combined ratio estimator.

$$\hat{X}_p^{(2)} = \frac{\hat{X}_p}{\hat{P}_p} P_p . \tag{3.2}$$

The post-stratified ratio estimator utilizes auxiliary information available at sub-class or post-stratum levels within the province. Let the subscripts {a} denote a collection of post-strata. The auxiliary information, namely population totals for each of the post-strata, denoted by P_{pa}, are available from outside sources. With $\hat{P}_{pa}$

and $\hat{X}_{pa}$ denoting the simple survey estimates of population and characteristic totals respectively, the post-stratified ratio estimator can be expressed as

$$\hat{X}_p^{(3)} = \sum_a \frac{\hat{X}_{pa}}{\hat{P}_{pa}} P_{pa} . \qquad (3.3)$$

For both the Labour Force Survey and the Survey of Consumer Finances, the official survey estimates are based on a post-stratified ratio estimation procedure. This gives rise to the sixth factor in the weighting of the survey data which, when multiplied by the subweight, yields the final weight for survey records. The post-stratum adjustment factor is the ratio of the external population estimate for the stratum divided by the simple survey estimated population total for the post-stratum, i.e. referring to (3.3), for post-stratum "a" the post-stratum adjustment factor is $P_{pa}/\hat{P}_{pa}$. For the Labour Force Survey, the post-strata are defined on the basis of age by sex groupings while for the Survey of Consumer Finances, the post-strata are defined as labour force status by class of worker groupings within each province. The outside estimates for the Survey of Consumer Finances are derived from Census and Labour Force Survey sources.

The third type of estimator is a generalized ratio estimator. Like the combined ratio estimator, this estimator utilizes information only at the provincial level, with an additional factor which takes account of the reliability of the simple survey estimates and the correlation between estimates of the population and the characteristic totals. As before, let $\hat{P}_p$ and $\hat{X}_p$ denote simple estimates of the total population and characteristic total respectively at the provincial level, and the total population count P_p is available from outside sources. The generalized ratio estimator is then defined as

$$\hat{X}_p^{(4)} = \frac{\hat{X}_p \; P_p}{\alpha \hat{P}_p + (1-\alpha) P_p} \tag{3.4}$$

where α is a suitably chosen constant. The optimum value of α, obtained by minimizing the mean square error of $\hat{X}_p^{(4)}$, is given by

$$\alpha_o = \frac{CV(\hat{X}_p)}{CV(\hat{P}_p)} \rho(\hat{X}_p, \hat{P}_p) \tag{3.5}$$

where $CV(\hat{X}_p)$ is the coefficient of variation of $\hat{X}_p$

$CV(\hat{P}_p)$ is the coefficient of variation of $\hat{P}_p$

and

$\rho(\hat{X}_p, \hat{P}_p)$ is the correlation coefficient between $\hat{X}_p$ and $\hat{P}_p$.

It may be noted that unlike the combined and post-stratified ratio estimators, this generalized ratio estimator (using α_o) is unbiased to the first degree of approximation and has the same mean square error as that of the regression estimator (see Singh [1969] and Naga Reddy [1974]). As this estimator has been found to be highly efficient in small empirical investigations, it is included in this empirical investigation to study if the gain in efficiency is substantial in the context of large scale surveys.

4. VARIANCE ESTIMATION

For both the surveys considered here, the methodology for estimating the sampling variability is modelled after the Keyfitz method. To apply this method of variance estimation, a set of pseudo-replicates within each stratum is required. For the purposes of variance estimation in the Labour Force Survey and the Survey of Consumer Finances in NSRU areas the n_h selected PSU's referred to as 'components' from stratum h are assumed to have been selected independently. In SRU areas, each subunit is split into two components for variance estimation purposes; this split is accomplished by dividing the selected clusters between the two components. In special areas, pseudo-strata are defined either as the design strata or as collapsed designed strata. Each pseudo-stratum is then sub-divided into from two to four components, the number depending on the sample sizes within each pseudo-stratum. We shall follow the notation presented in section 3 with the modification that the h denotes the strata or pseudo-strata and i denotes the components for variance estimation. The variance estimates for the estimators ($X^{(2)}$, $X^{(3)}$ and $X^{(4)}$) are based on the usual approximation to the variance of a ratio.

a) Variance estimate of the simple survey estimator:

$$\hat{Var}(\hat{X}_p) = \sum_{h\varepsilon p} \frac{n_h}{n_h-1} \sum_{i=1}^{n_h} (\hat{X}_{phi} - 1/n_h \sum_{i=1}^{n_h} \hat{X}_{phi})^2 \qquad (4.1)$$

where n_h is the number of components in stratum h.

b) Variance estimate of the combined ratio estimator:

Let $$\hat{D}_{phi} = \hat{X}_{phi} - \frac{\hat{X}_p}{\hat{P}_p} \hat{P}_{phi}. \qquad (4.2)$$

Then

$$\hat{Var}(\hat{X}_p^{(2)}) = (\frac{P_p}{\hat{P}_p})^2 \sum_{h\varepsilon p} \frac{n_h}{n_h-1} \sum_{i=1}^{n_h} [\hat{D}_{phi} - \frac{1}{n_h} \sum_{i=1}^{n_h} \hat{D}_{phi}]^2. \qquad (4.3)$$

c) Variance estimate of the post-stratified ratio estimator:

Let

$$\hat{D}_{phia} = \hat{X}_{phia} - \frac{\hat{X}_{pa}}{\hat{P}_{pa}} \hat{P}_{phia} . \tag{4.4}$$

Then

$$\hat{Var}(\hat{X}_p^{(3)}) = \sum_{h \epsilon p} \frac{n_h}{n_h - 1} \sum_{i=1}^{n_h} [\sum_a \frac{P_{pa}}{\hat{P}_{pa}} \hat{D}_{phia} - \frac{1}{n_h} \sum_{i=1}^{n_h} \sum_a \frac{P_{pa}}{\hat{P}_{pa}} \hat{D}_{phia}]^2 . \tag{4.5}$$

d) Variance estimate of the generalized ratio estimator:

Let

$$\hat{D}^*_{phi} = \hat{X}_{phi} - \frac{\hat{X}_p}{\alpha\hat{P}_p + (1-\alpha)P_p} \hat{P}_{phi} \tag{4.6}$$

Then

$$\hat{Var}(\hat{X}_p^{(4)}) = [\frac{\hat{P}_p}{\alpha\hat{P}_p + (1-\alpha)P_p}]^2 \sum_{h \epsilon p} \frac{n_h}{n_h - 1} \sum_{i=1}^{n_h} \{D^*_{phi} - \frac{1}{n_h} \sum_{i=1}^{n_h} D^*_{phi}\}^2 \tag{4.7}$$

It should be noted that although α is estimated from the sample, its effect on the sampling variability is neglected. A brief development of the variance expressions for the ratio estimators is presented in Appendix A.

5. EFFICIENCY COMPARISON

5.1 Formulation

The information supplied by an estimator is usually measured by the inverse of its variance (or its mean square error). Thus, the efficiency of an estimator t_2 over another estimator t_1 is generally defined as the ratio of the inverse of their variances, that is $E_{12} = V(t_1)/V(t_2)$. In actual practice, however, the true variances V(t) will not be available. One may compute the efficiency E.. by taking the ratio of the estimated variances v(t) or alternatively by taking the ratio of the estimated relative variances rv(t)= v(t)/t. The latter approach may be more suitable in situations where the estimates t. themselves vary significantly from one estimation method to another. We have chosen the latter approach for comparison of efficiencies. Thus, E_{ij} is the estimated efficiency of estimator t_j over another estimator t_i and is defined as the ratio of the inverse of their relative variances. In particular,

$$E_{12} = \frac{rv(\hat{X}^{(1)})}{rv(\hat{X}^{(2)})}$$ is the efficiency of the combined ratio estimator over the simple survey estimator.

The efficiency E_{12} may be expressed in terms of the correlation coefficient between $\hat{X}_p$ and $\hat{P}_p$, and is presented along with estimates for the Labour Force Survey in Appendix B.

$$E_{23} = \frac{rv(\hat{X}^{(2)})}{rv(\hat{X}^{(3)})}$$ is the efficiency of the post-stratified ratio estimator over the combined ratio estimator.

E_{23} may be thought of as a post-stratification index as it provides a measure of the gain in efficiency as a result of post-stratification. The overall efficiency of the post-stratified ratio estimator over the simple estimator is:

$$E_{13} = \frac{rv(\hat{X}^{(1)})}{rv(\hat{X}^{(3)})} .$$

Finally,

$$E_{24} = \frac{rv(\hat{X}^{(2)})}{rv(\hat{X}^{(4)})}$$

is the efficiency of the generalized ratio estimator over the usual combined ratio estimator.

It should be noted that $E_{13} = E_{12} E_{23}$ and $E_{14} = E_{12} E_{14}$ so that the overall efficiencies of the various ratio-type estimators relative to the simple survey estimator may be viewed as contributions by successive changes in the application of ratio adjustment procedures.

6. OBSERVATIONS

The following observations can be made, based on the results of Tables 1A, 1B and 1C.

1) The magnitude of gain in efficiency varies from characteristic to characteristic, depending mainly upon the degree of correlation between the characteristic and the total population. Maximum gain is achieved at the national level for Employed (having correlation about 0.92) followed by Not In Labour Force (having correlation about 0.67) and Unemployed which has approximate correlation 0.40. At the provincial level, the magnitude of the efficiencies differs from province to province but again a similar trend is noticeable for various characteristics examined.

2) Each of the ratio type estimators considered here is generally more efficient than the simple survey estimator.

3) The combined ratio estimator which is the simplest to compute achieves most of the efficiency gains for all the characteristics. For example, in case of Employed, the combined estimator is on an average five times more efficient than the simple estimator, (ranging from 3.65 to 6.66 times) whereas the additional gain in using post-stratified estimator is only about 50%. For Unemployed, however, the gain in using the combined estimator over the simple estimator is only about 25% and the additional gain in using either of the other two estimators is negligible.

4) The generalized ratio estimator is always more efficient than the combined ratio estimator. Although more than 30% gains are achieved for certain provinces for the characteristics Employed and Not in Labour Force, the gain at national level is only about 12%.

5) Post-stratified estimator, which is presently used in the Labour Force Survey, is slightly more efficient than the other two ratio type estimators. It may, however, be noted that this estimator uses more detailed data, i.e. data at the post-strata (age-sex) level whereas the other two ratio estimators use supplementary data only at the aggregate level.

6) In respect of estimates of Unemployed in Saskatchewan (Table 1B) both the combined and post-stratified estimators lead to a loss in efficiency as compared to the simple survey estimator, whereas the generalized ratio estimator even in this exceptional case remains slightly more efficient. This indicates that in situations where the unemployment level is exceptionally low the use of post-stratified estimator could lead to substantial loss in efficiency.

7) In addition to basing the above conclusions on a larger set of data, the primary reason for presenting three years data in Tables 1 (A,B,C) is to study the performance of these estimators over a period of time. It should be noted that the redesigned sample using 1971 Census and other more recent data became the official source of labour force estimates starting from January 1976. The results in Table 1A thus relate to the time when the sample was most up-to-date.

Comparing the values of E_{12} in Table 1A, based on 1976 with the corresponding values in Table 1B based on 1978, it is noticed that the gain in the relative efficiency of ratio estimator as compared to the simple estimator increased in 1978 for all provinces with the exception of Prince Edward Island. For the provinces of Alberta and Saskatchewan, this gain is two to three times the gain in efficiency achieved in 1976 and as a result there is a significant gain at the national level as well. It may thus be concluded that as we move away from the redesign time and the sample becomes more and more out-of-date, the simple survey estimator becomes less and less efficient and consequently the relative efficiency of the ratio estimators shows a gradual increase.

Further, it is noticed from Table 1C, based on 1979, that certain provinces show a decrease in the efficiency gain as compared to the gain in 1978 (Table 1B), and particularly so for the provinces of Alberta and Saskatchewan. This can be explained by the fact that a program of updating the sample was put in place during 1978/79 under which the samples from cities having larger growth are regularly updated and this program had maximum impact on the two provinces mentioned above.

It may, however, be noted that the additional gains of the other two ratio estimators remained relatively constant over time.

8) Until recently, only 18 age-sex groups (post-strata) were used in the LFS and therefore in Tables 1A, B, C, these groups were used for the empirical comparisons. Starting from 1979, there are now 40 age-sex groups used for the construction of post-stratified estimates. This increase was implemented primarily for the purpose of providing estimates for larger numbers or combinations of age groups. In table 2, a comparison is made between the two sets of post-strata. It is observed that the increase in the number of strata has no significant effect on the reliability of estimates. This, however, suggests that an optimum set of post-strata could be determined by repeating the study for various sets provided that corresponding population projections are available.

As noted in section 2, the Survey of Consumer Finances conducted in 1978 was a supplement to the LFS based on a sample of about 41,000 dwellings whereas the 1977 survey used a sample of about 17,000 dwellings selected independently from the LFS sampling frame. The estimated efficiencies for both survey years and especially for the 1977 survey, are subjected to larger sampling variability than the corresponding efficiencies as estimated for the Labour Force Survey data (particularly since the Labour Force Survey estimated efficiencies were averages over several monthly surveys). Despite the instability of the estimated relative efficiencies, the following patterns and relationships can be observed from Table 3 based on the 1977 and 1978 Surveys of Consumer Finances.

9) As in the case of labour force characteristics, the ratio estimators in this case as well, are more efficient than the simple survey estimator. The estimated gains in efficiency vary substantially from province to province. As well for some provinces there are large differences in the relative efficiencies between the two surveys.

10) Most of the gains in efficiency arise for the combined ratio estimator. The additional refinements due to the generalized ratio estimator and the post-stratified ratio estimator contribute relatively small gains in the efficiency.

11) The generalized ratio estimator is at least as efficient as the combined ratio estimator (See E 24) although the gain in efficiency over the combined ratio estimator varies considerably from province to province with observed gains at the national level of 6% and 11% for the 1977 and 1978 surveys respectively.

12) Although, at the national level, post-stratified ratio estimator results in a gain in efficiency as compared to the combined estimator (See E 23) (by 21% and 5% for 1977 and 1978 respectively), at the provincial level slight loss in relative efficiency is noticed for some province.

The post-stratified ratio estimator, however, ensures agreement of population totals between the survey and the control population totals derived from external sources (usually census data) within post-strata. This secondary benefit of the post-stratified ratio estimator is judged as being useful and important as it ensures a degree of consistency between the survey estimates and the external population totals down to the post-stratum level.

In summary, for both the Labour Force Survey and the Survey of Consumer Finances, considerable gains in efficiency, over the simple design based estimator, are realized by employing ratio-type estimation procedures. The combined ratio estimator provides most of the gains observed. The generalized ratio estimator generally provides minimal additional gains in efficiency over the combined ratio estimator, although for some provinces and for some characteristics, the gains approach 30-40%. The post-stratification in the majority of situations, provides marginal improvements in the efficiency of the post-stratified ratio estimator over the combined ratio estimator. For the Labour Force Survey, the comparison of efficiencies of the estimators over time yielded interesting results showing an increase in the performance of the ratio-type estimators as the sample design deteriorated.

It should be noted that all these comparisons are based on the estimated relative variances and to that extent the degree of confidence in the conclusions would depend upon the stability of these variance estimates. In the context of the LFS, the variance estimates should be highly stable due to the large size of the sample as well as due to the averaging of estimates over survey months. However, for the SCF, the variance estimates may not be as stable due to smallness of the sample size and data being based on single surveys, and therefore, studies using data from additional surveys would be needed to provide greater confidence in the results.

ACKNOWLEDGEMENT

The authors wish to thank the referee for some helpful comments.

Table 1A: Relative Efficiency of the Ratio Estimators*
Period: 1976 (January-June Average)

Provinces	Employed			Unemployed			Not In Labour Force		
	E_{12}	E_{23}	E_{13}	E_{12}	E_{23}	E_{13}	E_{12}	E_{23}	E_{13}
Nfld.	2.62	1.19	3.10	1.12	1.02	1.17	2.10	1.42	2.92
P.E.I.	5.38	2.22	12.18	0.96	0.88	0.86	2.28	1.59	3.10
N.S.	3.31	1.28	4.24	1.23	1.02	1.25	2.02	1.37	2.76
N.B.	2.19	1.74	3.80	1.02	0.98	1.00	1.80	1.82	3.24
Que.	3.24	1.44	4.62	1.02	1.02	1.04	2.40	1.49	3.53
Ont.	3.35	1.49	4.97	1.06	1.02	1.08	1.80	1.56	2.82
Man.	3.72	1.39	5.24	1.10	1.04	1.14	2.50	1.49	3.76
Sask.	6.40	1.74	11.02	1.17	0.98	1.17	1.85	2.28	4.16
Alta.	5.20	1.49	7.78	1.06	1.04	1.10	1.74	1.49	2.59
B.C.	4.54	2.96	6.25	1.23	1.08	1.32	1.74	1.51	2.62
Can.	3.65	1.46	5.34	1.08	1.02	1.10	2.02	1.56	3.13

* The efficiency of the generalized ratio estimator over the combined ratio estimator (i.e. E_{24}) is not available for this period.

Table 1B: Relative Efficiency of the Ratio Estimators
Period 1978 (January, April, July, October average)

Province	Employed				Unemployed				Not In Labour Force			
	E_{12}	E_{23}	E_{13}	E_{24}	E_{12}	E_{23}	E_{13}	E_{24}	E_{12}	E_{23}	E_{13}	E_{24}
Nfld.	4.49	1.12	5.11	2.10	1.51	0.98	1.46	1.02	1.96	1.21	2.34	2.07
P.E.I.	3.61	1.61	5.76	1.39	1.21	1.19	1.42	1.04	0.96	1.85	1.77	1.49
N.S.	3.02	1.37	4.00	1.04	1.12	1.04	1.17	1.02	1.88	1.51	2.82	1.04
N.B.	3.84	1.32	5.15	1.17	1.28	1.08	1.37	1.04	1.49	1.54	2.25	1.28
Que.	4.20	1.28	5.29	1.04	1.23	1.08	1.32	1.00	2.04	1.51	3.06	1.06
Ont.	4.97	1.61	11.16	1.12	1.25	1.06	1.32	1.08	1.49	1.93	2.82	1.17
Man.	4.16	1.46	6.10	1.14	1.04	1.02	1.06	1.00	1.69	1.72	2.76	1.14
Sask.	14.36	1.59	21.99	1.12	0.83	0.94	0.76	1.25	5.15	1.35	6.71	1.06
Alta.	17.81	1.88	35.52	1.39	1.42	0.96	1.35	1.04	3.39	1.90	6.66	1.35
B.C.	4.37	2.82	14.90	1.08	1.19	1.10	1.30	1.00	1.59	2.72	4.93	1.08
Canada	6.66	1.51	10.05	1.11	1.25	1.06	1.35	1.04	1.85	1.77	3.24	1.13

Table 1C: Relative Efficiency of the Ratio Estimators
Period: 1979 (January, April, July, October average)

Province	Employed				Unemployed				Not In Labour Force			
	E_{12}	E_{23}	E_{13}	E_{24}	E_{12}	E_{23}	E_{13}	E_{24}	E_{12}	E_{23}	E_{13}	E_{24}
Nfld.	3.04	1.33	3.97	1.04	1.23	1.08	1.32	1.02	2.60	1.34	3.98	1.06
P.E.I.	4.85	1.96	9.73	1.32	1.43	1.01	1.50	1.15	1.26	2.28	2.96	1.48
N.S.	4.38	1.51	6.37	1.17	1.19	1.06	1.26	1.05	1.69	1.65	2.80	1.19
N.B.	3.70	1.58	5.86	1.13	1.09	0.94	1.03	1.02	2.06	1.71	3.45	1.19
Que.	4.13	1.72	7.11	1.07	1.15	1.10	1.18	1.01	1.84	1.81	3.24	1.07
Ont.	7.28	1.56	10.64	1.07	1.30	1.06	1.37	1.03	1.71	1.65	2.81	1.11
Man.	5.84	1.91	11.17	1.21	1.15	1.04	1.21	1.07	1.22	1.88	2.71	1.20
Sask.	9.12	2.02	17.74	1.37	1.29	1.06	1.38	1.06	1.94	2.14	4.00	1.43
Alta.	9.84	1.68	15.40	1.10	1.16	1.01	1.17	1.04	1.71	1.87	3.19	1.10
B.C.	5.11	1.86	9.50	1.18	1.07	1.06	1.14	1.02	1.39	2.08	2.86	1.13
Canada	5.91	1.61	9.47	1.09	1.23	1.05	1.29	1.02	1.74	1.76	3.04	1.11

Table 2: Coefficients of Variation of LFS Estimates
(4 month average: Jan.79, April 79, July 79, Oct. 79)

Province	Employed		Unemployed	
	C.V.(40)	C.V.(18)	C.V.(40)	C.V.(18)
Nfld.	1.97	1.99	6.13	6.13
P.E.I.	2.12	2.15	8.67	8.79
N.S.	1.29	1.29	5.88	5.88
N.B.	1.37	1.37	5.37	5.37
Que.	0.75	0.75	3.68	3.67
Ont.	0.55	0.55	3.59	3.59
Man.	0.94	0.95	6.08	6.04
Sask.	0.94	0.96	7.80	7.78
Alta.	0.65	0.65	5.78	5.78
B.C.	0.87	0.87	4.74	4.80
Canada	0.32	0.32	1.87	1.87

Table 3: Relative Efficiency of Ratio Estimators*
Survey of Consumer Finances
Characteristic: Total Aggregate Income

Province	Period 1977, sample size = 17,000 approx.				Period 1978, sample size = 41,000 approx.			
	E_{12}	E_{23}	E_{13}	E_{24}	E_{12}	E_{23}	E_{13}	E_{24}
Nfld.	2.39	2.22	5.32	1.17	5.66	1.40	7.93	1.00
P.E.I.	2.11	0.88	1.86	1.00	2.67	2.13	5.70	1.42
N.S.	2.88	1.22	3.52	1.01	2.52	1.36	3.36	1.02
N.B.	5.02	1.79	8.98	1.04	3.55	1.60	5.67	1.15
Que.	5.54	1.44	7.98	1.06	3.57	1.09	3.88	1.07
Ont.	16.71	1.34	22.43	1.17	3.25	1.00	3.25	1.01
Man.	4.30	1.75	7.55	1.00	1.66	0.69	1.15	1.07
Sask.	4.50	1.45	6.55	1.00	11.22	0.94	10.53	1.42
Alta.	2.51	1.07	2.69	1.00	4.26	1.15	4.92	1.51
B.C.	2.63	1.53	4.04	1.00	6.42	0.97	6.44	1.01
Canada	9.76	1.21	11.81	1.06	3.80	1.05	3.99	1.11

* Labour force status x class of worker was considered as post-strata instead of age x sex groupings as in Table 1.

RESUME

Trois estimateurs par quotient sont considérés dans cet article. Il s'agit de l'estimateur combiné, de l'estimateur post-stratifié et d'un estimateur par quotient généralisé qui a déjà été proposé par Singh (1969) et par Naga Reddy (1974). Dans une évaluation empirique, on compare l'efficacité de ces estimateurs dans le contexte de deux enquêtes-ménages de grande envergure: l'Enquête sur la population active canadienne et l'Enquête sur les finances des consommateurs.

REFERENCES

[1] Keyfitz, V. (1957), "Estimates of Sampling Variance Where Two Units are Selected from Each Stratum", Journal of the American Statistical Association, 52,503-510.

[2] Cochran, W.G. (1963),"Sampling Techniques", 2nd Edition, John Wiley and Sons, New York.

[3] Reddy, V.N. (1974), "On a Transformed Ratio Method of Estimation", Sankhya, 36, 59-70.

[4] Singh, M.P. (1969), "Some Aspects of Estimation in Sampling from Finite Populations", Ph.D. Thesis (ch. 8), Indian Statistical Institute.

[5] Statistics Canada (1977), "Methodology of the Canadian Labour Force Survey" (1976), Catalogue No. 71-526.

[6] Turner, R. and Lawes, M. (1979), "Incomplete Data in the Survey of Consumer Finances", Statistics Canada (unpublished).

Appendix A

Considering the usual approximation for the variance of a ratio of statistics, the variance of the generalized ratio estimator can be expressed as:

$$\mathrm{Var}\ (X_p^{(4)}) = \mathrm{Var}\ \left(\frac{\hat{X}_p}{\alpha\hat{P}_p+(1-\alpha)P_p} P_p\right) \doteq P_p^2 \left\{\frac{E(\hat{X}_p)}{E(\alpha\hat{P}_p+(1-\alpha)P_p)}\right\}^2 \left\{ \frac{\mathrm{Var}(\hat{X}_p)}{[E(\hat{X}_p)]^2} - \right.$$

$$\left. - \frac{2\ \mathrm{Cov}(\hat{X}_p,\ \alpha\hat{P}_p + (1-\alpha)P_p)}{E(\hat{X}_p)\ E(\alpha\hat{P}_p + (1-\alpha)P_p)} + \frac{\mathrm{Var}(\hat{P}_p+1(1-\alpha)P_p)}{[E(\alpha\hat{P}_p+(1-\alpha)P_p)]^2}\right\}$$

$$= \left[\frac{P_p}{\alpha E(\hat{P}_p) + (1-\alpha)P_p}\right]^2 \left\{ \mathrm{Var}\ (\hat{X}_p) - 2\alpha \frac{E(\hat{X}_p)}{\alpha E(\hat{P}_p) + (1-\alpha)P_p} \mathrm{Cov}(\hat{X}_p,\hat{P}_p) \right.$$

$$\left. + \alpha^2 \left[\frac{E(\hat{X}_p)}{\alpha E(\hat{P}_p) + (1-\alpha)P_p}\right]^2 \mathrm{Var}(\hat{P}_p) \right\} .$$

Now using the inevitable procedure of estimating $E(\hat{X})$ and $E(\hat{P})$ by $\hat{X}$ and $\hat{P}$ respectively, $\mathrm{Var}\ (\hat{X}_p^{(4)})$ can be estimated by

$$\hat{\mathrm{Var}}\ (\hat{X}_p^{(4)}) = \left[\frac{P_p}{\alpha\hat{P}_p + (1-\alpha)P_p}\right]^2 \left\{ \hat{\mathrm{Var}}\ (\hat{X}_p)\ - 2\alpha \frac{\hat{X}_p}{\alpha\hat{P}_p + (1-\alpha)P_p}\ \hat{\mathrm{Cov}}\ (\hat{X}_p,\ \hat{P}_p) \right.$$

$$\left. + \alpha^2 \left[\frac{\hat{X}_p}{(\alpha\hat{P}_p + (1-\alpha)P_p}\right]^2 \hat{\mathrm{Var}}\ (\hat{P}_p) \right\}.$$

Now replacing $\hat{\mathrm{Var}}\ (\hat{X}_p)$ and $\hat{\mathrm{Var}}\ (\hat{P}_p)$ by expressions of the form (4.1) and with $\hat{\mathrm{Cov}}\ (\hat{X}_p,\ \hat{P}_p)$ estimated by

$$\hat{Cov}(\hat{X}_p,\hat{P}_p) = \sum_{h\varepsilon p} \frac{n_h}{h_h-1} \sum_{i=1}^{n_h} (\hat{X}_{phi} - 1/n_h \sum_{i=1}^{n_h} \hat{X}_{phi})(\hat{P}_{phi} - \frac{1}{n_h} \sum_{i=1}^{n_h} \hat{P}_{phi})$$

the above expression reduces to

$$[\frac{\hat{P}_p}{\alpha\hat{P}_p + (1-\alpha)P_p}]^2 \sum_{h\varepsilon p} \frac{n_h}{n_h-1} \sum_{i=1}^{n_h} \{\hat{X}_{phi} - \frac{\hat{X}_p}{\alpha\hat{P}_p + (1-\alpha)P_p} \hat{P}_{phi}$$

$$- \frac{1}{n_h} \sum_{i=1}^{n_h} (\hat{X}_{phi} - \frac{\hat{X}_p}{\alpha\hat{P}_p + (1-\alpha)P_p} \hat{P}_{phi})\}^2$$

$$\hat{Var}\ (\hat{X}_p^{(4)}) = [\frac{\hat{P}_p}{\alpha\hat{P}_p + (1-\alpha)P_p}]^2 \sum_{h\varepsilon p} \frac{n_h}{n_h-1} \sum_{i=1}^{n_h} \{D_{phi} - \frac{1}{n_h} \sum_{i=1}^{n_h} D_{phi}\}^2 \qquad (1)$$

$$\text{with } D_{phi} = \hat{X}_{phi} - \frac{\hat{X}_p}{\alpha\hat{P}_p + (1-\alpha)P_p} \hat{P}_{phi}\ . \qquad (2)$$

With $\alpha=1$ the generalized ratio estimator reduces to the usual ratio estimator, i.e. $\hat{X}_p^{(2)}$

and

$$Var\ (\hat{X}_p^{(2)}) = (\frac{P_p}{\hat{P}_p})^2 \sum_{h\varepsilon p} \frac{n_h}{n_h-1} \sum_{i=1}^{n_h} (D_{phi} - \frac{1}{n_h} \sum_{i=1}^{n_h} D_{phi})^2 \qquad (3)$$

$$\text{where } D_{phi} = \hat{X}_{phi} - \frac{\hat{X}_p}{\hat{P}_p} \hat{D}_{phi}\ . \qquad (4)$$

Similarly for the post-stratified ratio estimator, the variance can be expressed as

$$\mathrm{Var}\ (\hat{X}_p^{(3)}) = \mathrm{Var}\ (\sum_a \frac{\hat{X}_{pa}}{\hat{P}_{pa}} P_{pa})$$

$$= \sum_a \mathrm{Var}\ (\frac{\hat{X}_{pa}}{\hat{P}_{pa}} P_{pa}) + \sum_{a \neq b}\sum \mathrm{Cov}\ (\frac{\hat{X}_{pa}}{\hat{P}_{pa}} P_{pa}, \frac{\hat{X}_{pb}}{\hat{P}_{pb}} P_{pb}) . \quad (5)$$

Using expressions (3) and (4) $\mathrm{Var}\ (\frac{\hat{X}_{pa}}{\hat{P}_{pa}} P_{pa})$ is estimated by:

$$\hat{\mathrm{Var}}\ (\frac{\hat{X}_{pa}}{\hat{P}_{pa}} P_{pa}) = (\frac{P_{pa}}{\hat{P}_{pa}})^2 \sum_{h \varepsilon p} \frac{n_h}{n_h - 1} \sum_{i=1}^{n_h} (D_{phia} - \frac{1}{n_h} \sum_{i=1}^{n_h} D_{phia})^2 \quad (6)$$

$$\text{with } D_{phia} = \hat{X}_{phia} - \frac{\hat{X}_{pa}}{\hat{P}_{pa}} \hat{P}_{phia} . \quad (7)$$

Then for two different post-strata denoted by a and b,

$$\mathrm{Cov}\ (\frac{\hat{X}_{pa}}{\hat{P}_{pa}} P_{pa}, \frac{\hat{X}_{pb}}{\hat{P}_{pb}} P_{pb})$$

$$= P_{pa} P_{pb} \mathrm{Cov}\ (\frac{\hat{X}_{pa}}{\hat{P}_{pa}}, \frac{\hat{X}_{pb}}{\hat{P}_{pb}})$$

$$= P_{pa} P_{pb} \frac{E(\hat{X}_{pa})\ E(\hat{X}_{pb})}{E(\hat{P}_{pa})\ E(\hat{P}_{pb})} \{ \frac{\mathrm{Cov}\ (\hat{X}_{pa}, \hat{X}_{pb})}{E(\hat{X}_{pa})\ E(\hat{X}_{pb})} - \frac{\mathrm{Cov}\ (\hat{X}_{pa}, \hat{P}_{pb})}{E(\hat{X}_{pa})\ E(\hat{P}_{pb})}$$

$$- \frac{\mathrm{Cov}\ (\hat{X}_{pb}, \hat{P}_{pa})}{E(\hat{X}_{pb})\ E(\hat{P}_{pa})} + \frac{\mathrm{Cov}\ (\hat{P}_{pa}, \hat{P}_{pb})}{E(\hat{P}_{pa})\ E(\hat{P}_{pb})} \} \text{ up to 2nd order terms.}$$

Again estimating $E(\hat{X}_{pa})$, $E(\hat{X}_{pb})$, $E(\hat{P}_{pa})$ and $E(\hat{P}_{pb})$ by $\hat{X}_{pa}$, $\hat{X}_{pb}$, $\hat{P}_{pa}$ and $\hat{P}_{pb}$ respectively,

$$\hat{Cov}\left(\frac{\hat{X}_{pa}}{\hat{P}_{pa}} P_{pa}, \frac{\hat{X}_{pb}}{\hat{P}_{pb}} P_{pb}\right) = \frac{P_{pa}}{\hat{P}_{pa}} \frac{P_{pb}}{\hat{P}_{pb}} [Cov(\hat{X}_{pa}, \hat{X}_{pb}) - \frac{\hat{X}_{pb}}{\hat{P}_{pb}} Cov(\hat{X}_{pa}, \hat{P}_{pb})$$

$$- \frac{\hat{X}_{pa}}{\hat{P}_{pa}} Cov(\hat{X}_{pb}, \hat{P}_{pa}) + \frac{\hat{X}_{pa}}{\hat{P}_{pa}} \frac{\hat{X}_{pb}}{\hat{P}_{pb}} Cov(\hat{P}_{pa}, \hat{P}_{pb})]$$

$$= \frac{P_{pa}}{\hat{P}_{pa}} \frac{P_{pb}}{\hat{P}_{pb}} [\sum_{h\varepsilon p} \frac{n_h}{n_h - 1} \sum_{i=1}^{n_h} (\hat{X}_{phia} - \frac{1}{n_h} \sum_{i=1}^{n_h} \hat{X}_{phia})(\hat{X}_{phib} - \frac{1}{n_h} \sum_{i=1}^{n_h} \hat{X}_{phib})$$

$$- \frac{\hat{X}_{pb}}{\hat{P}_{pb}} \sum_{h\varepsilon p} \frac{n_h}{n_h - 1} \sum_{i=1}^{n_h} (\hat{X}_{phia} - \frac{1}{n_h} \sum_{i=1}^{n_h} \hat{X}_{phia})(\hat{P}_{phib} - \frac{1}{n_h} \sum_{i=1}^{n_h} \hat{P}_{phib})$$

$$- \frac{\hat{X}_{pa}}{\hat{P}_{pa}} \sum_{h\varepsilon p} \frac{n_h}{n_h - 1} \sum_{i=1}^{n_h} (\hat{X}_{phib} - \frac{1}{n_h} \sum_{i=1}^{n_h} \hat{X}_{phib})(\hat{P}_{phia} - \frac{1}{n_h} \sum_{i=1}^{n_h} \hat{P}_{phia})$$

$$+ \frac{\hat{X}_{pa}}{\hat{P}_{pa}} \frac{\hat{X}_{pb}}{\hat{P}_{pb}} \sum_{h\varepsilon p} \frac{n_h}{n_h - 1} \sum_{i=1}^{n_h} (\hat{P}_{phia} - \frac{1}{n_h} \sum_{i=1}^{n_h} \hat{P}_{phia})(\hat{P}_{phib} - \frac{1}{n_h} \sum_{i=1}^{n_h} \hat{P}_{phib})]$$

$$= \frac{P_{pa}}{\hat{P}_{pa}} \frac{P_{pb}}{\hat{P}_{pb}} \sum_{h\varepsilon p} \frac{n_h}{n_h - 1} \sum_{i=1}^{n_h} [\{(\hat{X}_{phia} - \frac{\hat{X}_{pa}}{\hat{P}_{pa}} \hat{P}_{phia}) - \frac{1}{n_h} \sum_{i=1}^{n_h} (\hat{X}_{phia} - \frac{\hat{X}_{pa}}{\hat{P}_{pa}} \hat{P}_{phia})\}$$

$$- \{(\hat{X}_{phib} - \frac{\hat{X}_{pb}}{\hat{P}_{pb}} \hat{P}_{phib}) - \frac{1}{n_h} \sum_{i=1}^{n_h} (\hat{X}_{phib} - \frac{\hat{X}_{pb}}{\hat{P}_{pb}} \hat{P}_{phib})\}]^2$$

$$= \frac{P_{pa}}{\hat{P}_{pa}} \frac{P_{pb}}{\hat{P}_{pb}} \sum_{h \epsilon p} \frac{n_h}{n_h - 1} \sum_{i=1}^{n_h} [\{D_{phia} - \frac{1}{n_h} \sum_{i=1}^{n_h} D_{phia}\}\{D_{phib} - \frac{1}{n_h} \sum_{i=1}^{n_h} D_{phib}\}]^2 .$$

Now equation (5) is estimated by :

$$\hat{Var}\ (\hat{X}_p^{(3)}) = \sum_{a} \left(\frac{P_{pa}}{\hat{P}_{pa}} \right)^2 \sum_{h \epsilon p} \frac{n_h}{n_h - 1} \sum_{i=1}^{n_h} (D_{phia} - \frac{1}{n_h} \sum_{i=1}^{n_h} D_{phia})^2$$

$$+ \sum_{a \neq b} \sum \frac{P_{pa}}{\hat{P}_{pa}} \frac{P_{pb}}{\hat{P}_{pb}} \sum_{h \epsilon p} \frac{n_h}{n_h - 1} \sum_{i=1}^{n_h} [\{D_{phia} - \frac{1}{n_h} \sum_{i=1}^{n_h} D_{phia}\}\{D_{phib} - \frac{1}{n_h} \sum_{i=1}^{n_h} D_{phib}\}]$$

$$= \sum_{h \epsilon p} \frac{n_h}{n_h - 1} \sum_{i=1}^{n_h} [\sum_{a} \frac{P_{pa}}{\hat{P}_{pa}} D_{phia} - \frac{1}{n_h} \sum_{i=1}^{n_h} \sum_{a} \frac{P_{pa}}{\hat{P}_{pa}} D_{phia}]^2$$

which is equivalent to expression (4.5) in the body of the paper.

Appendix B

The efficiency of the combined ratio estimator versus the simple survey estimator can be derived and expressed in terms of the correlation coefficient between the numerator and denominator of the ratio of characteristic to population totals (see for example Cochran, Section 6.8, p. 165). Specifically,

$$E_{12} = \{1 - 2 \frac{\sqrt{rv(\hat{P}_p)}}{\sqrt{rv(\hat{X}_p)}} [\hat{\rho}(\hat{X}_p, \hat{P}_p) - \frac{1}{2} \frac{\sqrt{rv(\hat{P}_p)}}{\sqrt{rv(\hat{X}_p)}}]\}^{-1}$$

where $\hat{\rho}(\hat{X}_p, \hat{P}_p)$ denotes the estimated correlation coefficient between the simple survey estimates $\hat{X}_p$ and $\hat{P}_p$ and rv denotes the relative variance.

The following table presents the estimated relative efficiencies E_{12} and estimated correlation coefficients $\hat{\rho}$ for the Labour Force Survey for the three periods 1976, 1978 and 1979 considered in this report. (The values E_{12} and $\hat{\rho}$ are averages of the corresponding monthly estimates).

Table 4: Estimated Relative Efficiencies and Correlation Coefficients at the Canada Level.

Characteristic

Period	EMPLOYED		UNEMPLOYED		NOT IN LABOUR FORCE	
	E_{12}	$\hat{\rho}$	E_{12}	$\hat{\rho}$	E_{12}	$\hat{\rho}$
1976	3.65	0.853	1.08	0.245	2.02	0.708
1978	6.66	0.925	1.25	0.441	1.85	0.703
1979	5.91	0.915	1.23	0.318	1.74	0.676