EVALUATION OF COMPOSITE ESTIMATION FOR THE CANADIAN LABOUR FORCE SURVEY[1]

S. Kumar and H. Lee[2]

This study considers the suitability of composite estimation techniques for the Canadian Labour Force Survey. The performance of a class of AK composite estimators introduced initially by Gurney and Daly is investigated for several characteristics. While the ordinary composite estimate has a large bias, the AK composite estimate is capable of reducing the bias. Composite estimates having minimum variance and minimum mean square error are compared.

1. INTRODUCTION

The Canadian Labour Force Survey (LFS) is conducted each month by Statistics Canada and is designed to produce estimates for various labour force characteristics. The LFS sample design follows a rotation scheme that permits the replacement of one-sixth of the households in the sample each month (see [7]). The sample is composed of six panels or rotation groups. A panel remains in the sample for a period of six consecutive months.

As pointed out in Bailar [1], one of the major drawbacks of composite estimation currently in use for the U.S. Current Population Survey (CPS) is its bias as compared to the simple ratio estimator for estimates of level. This bias stems from rotation group differences: the phenomenon that estimates based on data from different panels relating to the same time period do not have the same expected value. This phenomenon, often referred to as the rotation group bias, has been studied for LFS (see [2] and [6]). Recently, Huang and Ernst [4] have reported results in the context of the CPS on the performance of AK composite estimator introduced initially by Gurney and Daly [3]. A and K are

---

[1] Presented at the Joint Statistical Meetings of the American Statistical Association, the Biometric Society, the Institute of Mathematical Statistics and the Statistical Society of Canada in Toronto, August 1983.

[2] S. Kumar and H. Lee, Census and Household Survey Methods Division, Statistics Canada.

constants in the equation defining the composite estimator. Their results show improvement over the composite estimates currently in use for CPS as regards variance and bias.

The objective of this investigation is to study the suitability of composite estimation techniques for LFS. In this study the performance of different composite estimators of level and change will be investigated for the following five characteristics; in labour force, employed, employed agriculture, employed non-agriculture, and unemployed. These composite estimators are compared with the simple ratio estimator which is presently in use for LFS. The study is based on the province of Ontario data for 1980-81.

2. DEFINITIONS AND NOTATION

We are interested in estimating Y_m the number of persons in the population with a certain characteristic for the month m. Let

$y_{m,i}$ = A simple ratio estimator of Y_m based on the i-th panel ($i = 1,2,\ldots,6$). Here the i-th panel refers to the sub-sample (rotation group) that is in the sample for the i-th time. It will be referred to as the i-th panel estimator.

$d_{m,m-1}$ = estimator of change $(Y_m - Y_{m-1})$ from the month $(m - 1)$ to the month m based on five panels that are common to the months m and $(m - 1)$

$$= \sum_{j=2}^{6} (y_{m,j} - y_{m-1,j-1})/5. \tag{2.1}$$

y'_m = AK composite estimator of Y_m defined as

$$y'_m = (1 - K + A)y_{m,1}/6 + (1 - K - \frac{A}{5}) \sum_{j=2}^{6} y_{m,j}/6$$

$$+ K(y'_{m-1} + d_{m,m-1}) \tag{2.2}$$

where K and A are constants, and $0 \leq K < 1$.

The equation (2.2) defines a class of estimators referred to as AK composite estimators. The estimators obtained by taking $A = 0$ in (2.2) are referred to as K composite estimators. The simple ratio estimator, to be denoted by $\bar{y}_m$, the mean of six panel estimators can be obtained by taking $A = 0$ and $K = 0$ in (2.2). We investigate the relative performance of the optimal (minimum variance or minimum mean square error) AK composite, K composite and simple ratio estimators.

We assume the rotation group bias $E(y_{m,i}) - Y_m$ is independent of m and is a function of i. We denote this bias by α_i. Formally

$$\alpha_i = E(y_{m,i}) - Y_m. \tag{2.3}$$

The expression for the bias of the composite estimator is given in Appendix I.

3. ASSUMPTIONS

The rotation system in the LFS is schematically described in Table 1, where the current (month m) panel i (= 1,2,...,6, denoting interview month no.) is the same as panel i - j in month m - j, provided i - j lies between 1 and 5. The immediate predecessor to panel i of month m as of month m - j is given by (6 + i - j) provided (6 + i - j) lies between 1 and 6. Likewise, the second predecessor to panel i as of month m - j is given by (12 + i - j) provided (12 + i - j) lies between 1 and 6. In general, the r-th predecessor to panel i of month m is given by (i - j + 6r) in month m - j. Note that the 0-th predecessor to a panel means the same panel in earlier months.

The expression for the variance of y'_m, i.e. $V(y'_m)$ involves the variances and covariances of various panel estimators (see Appendix II). The following variance-covariance structure for various panel estimators is assumed. The assumptions conform to the LFS rotation pattern, illustrated in Table 1.

(i) $V(y_{m,i}) = \sigma^2$ for all m and $i = 1,2,\ldots,6$,

(ii) $Cov(y_{m,i}, y_{m-j,i-j+6r}) = \gamma_j^{(r)}\sigma^2$, where $i = 1,2,\ldots,6$, $j > 0$ and $r \geq 0$, such that $6 \geq i - j + 6r \geq 1$. Here r denotes the number of predecessors to the current panel.

For $r = 0$, i.e., $6 > i - j \geq 1$, let $\gamma_j^{(r)} = \rho_j$ (based on overlapping panels of months m and m - j).

For $r = 1$, i.e., $6 \geq i - j + 6 \geq 1$, let $\gamma_j^{(r)} = \gamma_j$ (based on the current panel and its immediate predecessor j months back).

For $r \geq 2$, i.e., $6 \geq i - j + 6r \geq 1$, let $\gamma_j^{(r)} = 0$ (based on the current panel and its r-th predecessor j months back).

(iii) Of interest to the development of the variance of the composite estimator y_m' are the correlation coefficients ρ_j and γ_j, both of which are assumed to be stationary; i.e. they are functions of j and not of m. It is reasonable to assume that both ρ_j's and γ_j's are positive since ρ_j's are based on characteristics of largely common households while γ_j's are based on the characteristics of households in the current month and those of their near (in many cases next door) neighbours j months back (apart from cluster rotation).

(iv) The expression for $V(y_m')$ contains covariance terms not included in the assumptions (ii) and (iii). Some of these are:

$Cov(y_{m,i}, y_{m,j})$ for $i \neq j$, $Cov(y_{m,i}, y_{m-1,j})$ for $i = 1$, $j \neq 6$, and $i \neq 1$, $j \neq i - 1$, and $Cov(y_{m,i}, y_{m-g,j})$ for $g \geq 12$. These and all other covariances not defined above, including those with $\gamma_j^{(2)}$ and existing in the expression for $V(y_m')$ are assumed to be zero.

Following these assumptions, a variance expression for the AK composite estimator was derived in terms of the above parameters. The mathematical details for derivation of the expression for the bias and variance of y'_m, and the variance of $y'_m - y'_{m-1}$ are given in the appendices.

4. RESULTS AND DISCUSSION

The quantities σ^2, ρ_j and γ_j in the expresssion for $V(y'_m)$ were replaced by their estimates (For details of the methodology for estimating ρ's and γ's, see [5]). Note that, in the Canadian LFS ρ_j's do not exist for $j \geq 6$ because of no overlapping panels. Nor do γ_j's exist for $j \geq 12$ because for $j \geq 12$, there exist 2nd or higher order predecessors to the current panel and the correlation may be taken as 0 in the developments. Estimates of ρ_j, $\hat{\rho}_j$, are given in Table 2. The estimate of ρ_5 has been obtained by extrapolating other ρ_j's as it was not possible to estimate it directly from the sample. Note that $\hat{\rho}_j$ $(j = 1,2,\ldots,5)$ is a decreasing function of j for all the five characteristics. This is consistent with what we expect intuitively about the behaviour of ρ_j's. Also ρ_j's are high for all the characteristics except "unemployed".

Table 3 gives the estimates $\hat{\gamma}_j$ of γ_j. The estimates $\hat{\gamma}_5$ and $\hat{\gamma}_{11}$ were obtained respectively by interpolating and extrapolating other $\hat{\gamma}_j$'s. Intuitively, we expect γ_j's to decrease with j for each characteristic. We observe that this is not the case with $\hat{\gamma}_j$'s. Although $\hat{\gamma}_j$'s do not exhibit monotonic decreasing behaviour, we point out that whenever the difference $(\hat{\gamma}_{j+1} - \hat{\gamma}_j)$ is positive, its magnitude is very small. The positiveness of these differences could be due to the sampling variability rather than a real positiveness of $(\gamma_{j+1} - \gamma_j)$.

In the following discussion, the term relative efficiency of AK composite (or K composite) estimator refers to its efficiency relative to the simple ratio estimator.

Tables 4A and 4B give the results of comparing the estimated variances of three estimators. These are: (i) optimal AK composite estimator, i.e., an estimator having minimum variance among the class of estimators defined by (2.2), (ii) optimal K composite estimator (obtained by taking A = 0 in (2.2) and having minimum variance among all estimators in this subclass), and (iii) the simple ratio estimator. For $0 \leq K < 1$, nearly optimal values of K and (K, A) are also given (K was incremented by 0.1 and the optimal value of A was determined for each fixed K). Here, a value (K, A) is referred to optimal value if the AK composite estimator with this value has the smallest variance among all AK composite estimators defined by (2.2). Similar definition applies to the term "optimal K". Table 4A (computed using $\hat{\gamma}_j$'s given in Table 3) shows that, for all characteristics except "unemployed" there are 18-21% gains in relative efficiency for the K composite estimates and 26-30% gains in the relative efficiency for the AK composite estimates.

To determine the effect of γ_j's on the relative efficiencies, γ_j's were replaced by zero's in the expression for $V(y_m')$ and the optimal K, optimal (K, A), and the relative efficiencies were computed. These results are presented in Table 4B. Note that the optimal K's and optimal (K, A)'s in the Tables 4A and 4B are different. Comparison of the corresponding relative efficiencies in these two tables shows that positive γ's have a negative effect on the reduction in variance, i.e., the gains in relative efficiency are reduced. The greatest reduction in relative efficiency is for the characteristic "employed agriculture". This is the characteristic with relatively high values of $\hat{\gamma}_j$'s. Thus taking γ_j's to be zero, when $\gamma_j > 0$, can result in over-estimation of the relative efficiencies and the degree of over-estimation depends on the magnitude of γ_j's.

As mentioned in the introduction, one of the drawbacks of the composite estimators of level is their bias as compared to the simple ratio estimator. Thus comparing the variances of biased estimators can sometimes result in erroneous conclusions about the relative performance of these estimators. It is appropriate to examine the mean square error in the case of biased estimators. The expression for the bias of y_m' (see Appendix I) involves α_i's (the rotation group biases). The quantity $\hat{\alpha}_i = y_{m,i} - \hat{Y}_m$ is an unbiased estimator of α_i if $\hat{Y}_m$ is an unbiased estimator of Y_m. We assume that the simple ratio estimator $\bar{y}_m$ is an unbiased estimator of Y_m, i.e., $\sum_{i=1}^{6} \alpha_i = 0$. Values of $\hat{\alpha}_i$ ($i = 1,2,...,6$) for various characteristics are given in Table 5. For each of three characteristics "in labour force", "employed" and "employed non-agriculture", we note that: (i) $\hat{\alpha}_1$ is negative while all other $\hat{\alpha}_i$'s are positive; and (ii) $\hat{\alpha}_1$ is large relative to the other $\hat{\alpha}_i$'s.

Table 6 gives the values of optimal K, the optimal (K, A) and results of comparing mean square errors. The optimal K was determined among 10 values of K = 0(0.1)0.9 in the same manner for Tables 4A and 4B. However, the optimal (K, A) was computed in a different way. It was chosen among all possible combinations of K = 0(0.1)0.9 and A = 0(0.1)1.0 rather than determining optimal A for each fixed K = 0(0.1)0.9 (as used for Tables 4A and 4B). Two criteria of optimality are used. One is based on the concept of minimum variance (as is the case for Tables 4A and 4B), and the other is based on the concept of minimum mean square error.

It is shown in Appendix I that

$$E(y'_m) = Y_m + [A\alpha_1 + K(\alpha_6 - \alpha_1)]/[5(1-K)].$$

Bias of each estimate in Table 6 is computed by using $\hat{\alpha}_1$ and $\hat{\alpha}_6$ (given in Table 5) instead of α_1 and α_6 in the above formula. Now we discuss the results of Table 6.

For the K composite estimate (based on minimum mean square error optimality) there is only a moderate gain in relative efficiency for the characteristic "employed agriculture" and a nominal gain for the characteristic "unemployed". Also, the bias of the estimates for these two characteristics is small. For the remaining characteristics, the simple ratio estimate is the optimal K composite estimate.

The K composite estimates (considered in Table 4A and based on minimum variance optimality) for the three characteristics "in labour force", "employed" and "employed non-agriculture" have relative efficiencies less than 10%. In these cases, the poor performance can be attributed to the large bias. For each of the remaining two characteristics, K composite estimate is only marginally better than the simple ratio estimate, i.e., the gain in relative efficiency is insignificant. The difference in the corresponding relative efficiency results in Tables 4A and 6 is due to the different relative efficiency definitions used for the two tables. For Table 4A, relative efficiency is defined as the ratio of appropriate variances whereas for Table 6, mean square errors are used instead of the variances.

The AK composite estimate (based on minimum mean square optimality) shows relative efficiency gains in the range 16-22% for all characteristics except "unemployed". Also, the bias of estimate for each characteristic is small.

However, the AK composite estimate based on minimum variance optimality, like the corresponding K composite estimate, has very low relative efficiency for the characteristics "in labour force", "employed", "employed non-agriculture" because of large bias in these cases. The gain in relative efficiency for the characteristic "employed agriculture" is moderate whereas the corresponding gain the characteristic "unemployed" is nominal.

The results in Table 6 show that, among the four composite estimates discussed above, the optimal AK composite estimates (based on minimum mean square error) have relative efficiencies higher for all characteristics than the corresponding relative efficiencies for other composite estimates. We will discuss later the results in the last column of Table 6.

We note, from the expression for $E(y_m')$ given earlier, the $y_m' - y_{m-1}'$ is an unbiased estimator of $Y_m - Y_{m-1}$, i.e., K or AK composite estimators of change are unbiased. Table 7 gives the optimal K, optimal (K, A), and relative efficiency results for optimal K composite and optimal AK composite estimates of change. The gains in relative efficiency for the characteristics "in labour force", "employed", and "employed non-agriculture" are in the 46-55%

range for K composite and AK composite estimates. For the characteristic "employed agriculture", the optimal AK composite estimate is also optimal K composite and the gain in relative efficiency is about 135%. The gain in relative efficiency for the characteristic "unemployed" is about 6% for both estimates.

It should be pointed out that the optimal value of K or (K, A) is characteristic dependent. Thus the additive property of the estimates is not preserved when different values of K or (K, A) are employed. To preserve additivity, a common value of $K = 0.4$ and $A = 0.4$ was selected for estimates of level and change. The following remarks describe the performance of the AK composite estimate with $K = 0.4$ and $A = 0.4$. The last column of Table 6 shows that the gains in relative efficiency for AK composite estimates of level are in the 6-10% range for all characteristics except "unemployed". The results of Table 7 show that the gains in relative efficiency for AK composite estimates of change are in the 12-15% range for all characteristics except "unemployed". The gain in relative efficiency for AK composite estimates of level and change is about 2-3% for the characteristic "unemployed".

ACKNOWLEDGEMENT

The authors wish to thank D. Drew, G.B. Gray and M.P. Singh of Statistics Canada for helpful discussions during the progress of this work, and the referee whose suggestions led to many improvements.

TABLE 1

Common and Predecessor Panels Pertaining
To Months m and m-j

Panels in Month m	m-1	m-2	m-3	m-4	m-5	m-6	m-7	m-8	m-9	m-10	m-11
1	(6)	(5)	(4)	(3)	(2)	(1)	((6))	((5))	((4))	((3))	((2))
2	1	(6)	(5)	(4)	(3)	(2)	(1)	((6))	((5))	((4))	((3))
3	2	1	(6)	(5)	(4)	(3)	(2)	(1)	((6))	((5))	((4))
4	3	2	1	(6)	(5)	(4)	(3)	(2)	(1)	((6))	((5))
5	4	3	2	1	(6)	(5)	(4)	(3)	(2)	(1)	((6))
6	5	4	3	2	1	(6)	(5)	(4)	(3)	(2)	(1)

The correlation coefficients between common panels of months m and m-j indicated by panels with no parentheses equal ρ_j.

The correlation coefficient between panels of month m and their "single" predecessor of month m-j equals γ_j, the panels indicated by single parentheses.

The correlation coefficient between panels of month m and their double predecessor of month m-j equals $\gamma_j^{(2)}$, the panels indicated by double parentheses. In this report, all $\gamma_j^{(2)}$ are assumed to equal 0.

TABLE 2

Estimated Correlation ρ's (1980-1981 Ontario)

Characteristics \ $\hat{\rho}$	$\hat{\rho}_1$	$\hat{\rho}_2$	$\hat{\rho}_3$	$\hat{\rho}_4$	$\hat{\rho}_5$
In Labour Force	.843	.782	.717	.674	.631
Employed	.852	.779	.709	.664	.619
Employed Agriculture	.955	.926	.901	.861	.821
Employed Non-Agriculture	.861	.791	.724	.678	.632
Unemployed	.580	.445	.334	.286	.238

TABLE 3

Estimated Correlation γ's (1980-1981 Ontario)

Characteristics \ $\hat{\gamma}$	$\hat{\gamma}_1$	$\hat{\gamma}_2$	$\hat{\gamma}_3$	$\hat{\gamma}_4$	$\hat{\gamma}_5$	$\hat{\gamma}_6$	$\hat{\gamma}_7$	$\hat{\gamma}_8$	$\hat{\gamma}_9$	$\hat{\gamma}_{10}$	$\hat{\gamma}_{11}$
In Labour Force	.161	.141	.128	.133	.135	.136	.125	.127	.124	.122	.127
Employed	.164	.136	.142	.142	.146	.149	.148	.150	.153	.141	.148
Employed Agriculture	.477	.483	.474	.486	.480	.474	.459	.429	.394	.323	.252
Employed Non-Agriculture	.184	.150	.147	.157	.162	.167	.166	.169	.174	.156	.166
Unemployed	.141	.074	.076	.063	.057	.051	.045	.060	.077	.136	.074

TABLE 4A & 4B

The Optimal (K, A) and K, and the Relative Efficiencies of K Composite and AK Composite Estimators.

TABLE 4A

$\gamma_i \neq 0$

Characteristics	K composite		AK composite		
	Optimal K	Relative Efficiency	Optimal K	Optimal A	Relative Efficiency
In Labour Force	0.7	118.8	0.8	0.48	128.4
Employed	0.7	118.5	0.8	0.49	128.1
Employed Agriculture	0.8	120.6	0.8	0.38	126.9
Employed Non-Agriculture	0.7	119.4	0.8	0.47	129.3
Unemployed	0.3	102.8	0.5	0.38	105.2

TABLE 4B

$\gamma_i = 0$ for all i.

Characteristics	K composite		AK composite		
	Optimal K	Relative Efficiency	Optimal K	Optimal A	Relative Efficiency
In Labour Force	0.7	125.5	0.8	0.50	138.4
Employed	0.7	125.3	0.8	0.51	137.9
Employed Agriculture	0.8	167.3	0.9	0.46	187.9
Employed Non-Agriculture	0.7	126.9	0.8	0.49	140.2
Unemployed	0.4	104.4	0.6	0.51	108.4

Relative efficiency is with respect to the simple ratio estimator and is defined as 100 times the ratio V(simple ratio estimator)/V(K or AK composite).

TABLE 5

Estimates (in thousands) of Rotation Group Bias α_i

Characteristics	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\alpha}_3$	$\hat{\alpha}_4$	$\hat{\alpha}_5$	$\hat{\alpha}_6$
In Labour Force	-135.3	39.8	41.1	31.1	15.4	7.9
Employed	-141.7	35.5	34.9	31.3	25.4	14.8
Employed Agriculture	-4.2	-2.6	2.2	-0.1	4.2	0.5
Employed Non-Agriculture	-137.5	38.0	32.7	31.3	21.2	14.3
Unemployed	6.4	4.3	6.2	-0.1	-9.9	-6.9

α_i is defined as $E(y_{m,i}) - Y_m$ and estimated by $y_{m,i} - \sum_{i=1}^{6} y_{m,i}/6$.

TABLE 6

Comparison of the Variances and the Mean Square Errors of Simple, K Composite, and AK Composite Estimators

IN LABOUR FORCE

	Simple Ratio Estimator	K Composite (Min MSE) K=0	K Composite (Min Var) K=0.7	AK Composite (Min MSE) K=0.7 A=0.7	AK Composite (Min Var) K=0.8 A=0.5	Common K, A for all Characteristics K=0.4 A=0.4
Monthly Level Estimate 10^3	4480.7	4480.7	4547.5	4484.4	4527.6	4481.7
Variance 10^6	432.0	432.0	363.8	358.1	336.5	391.8
Bias 10^3	0	0	66.8	3.7	46.9	1.0
Mean Square Error 10^6	432.0	432.0	4284.5	371.5	2532.4	392.9
Relative Efficiency		100.0	9.0	116.3	17.1	110.0

EMPLOYED

		K=0	K=0.7	K=0.8 A=0.9	K=0.8 A=0.5	K=0.4 A=0.4
Monthly Level Estimate 10^3	4186.0	4186.0	4259.0	4183.6	4240.3	4188.0
Variance 10^6	473.3	473.3	399.6	397.7	369.5	428.9
Bias 10^3	0	0	73.0	-2.4	54.3	2.0
Mean Square Error 10^6	473.3	473.3	5732.2	403.2	3320.9	432.8
Relative Efficiency		100.0	8.3	117.4	14.3	109.4

EMPLOYED AGRICULTURE (TABLE 6 continued)

		Simple Ratio Estimator	K Composite		AK Composite		Common K, A for all Characteristics
			(Min MSE) K=0.6	(Min Var) K=0.8	(Min MSE) K=0.8 A=0.6	(Min Var) K=0.8 A=0.4	K=0.4 A=0.4
Monthly Level Estimate	10^3	142.0	143.4	145.7	143.2	144.1	142.1
Variance	10^6	85.7	75.6	71.1	68.7	67.6	80.8
Bias	10^3	0	1.4	3.7	1.2	2.1	0.1
Mean Square Error	10^6	85.7	77.6	85.1	70.2	71.8	80.8
Relative Efficiency			110.5	110.7	122.2	119.4	106.1

EMPLOYED NON-AGRICULTURE

			K=0	K=0.7	K=0.8 A=0.9	K=0.8 A=0.5	K=0.4 A=0.4
Monthly Level Estimate	10^3	4043.9	4043.9	4114.7	4041.6	4096.6	4045.8
Variance	10^6	498.9	498.9	417.8	418.0	385.9	452.8
Bias	10^3	0	0	70.8	-2.3	52.7	1.9
Mean Square Error	10^6	498.9	498.9	5436.1	423.3	3161.7	456.4
Relative Efficiency			100.0	9.2	117.9	15.8	109.3

UNEMPLOYED (TABLE 6 continued)

	Simple Ratio Estimator	K Composite (Min MSE) K=0.2	K Composite (Min Var) K=0.3	AK Composite (Min MSE) K=0.4 A=0.4	AK Composite (Min Var) K=0.5 A=0.4	Common K, A for all Characteristics K=0.4 A=0.4
Monthly Level Estimate 10^3	294.8	294.1	293.7	293.9	293.2	293.9
Variance 10^6	117.5	114.9	414.3	112.5	111.7	112.5
Bias 10^3	0	-0.7	-1.1	-0.9	-1.6	-0.9
Mean Square Error 10^6	117.5	115.4	115.7	113.3	114.4	113.3
Relative Efficiency		101.9	101.6	103.7	102.7	103.7

Relative efficiency is relative to the simple ratio estimator and is defined by 100 times MSE(simple ratio estimator)/MSE(K or AK Composite estimator).

TABLE 7

Relative Efficiency of Composite Estimators
for Month-to-Month Change

Labour Force Characteristics	K composite		AK Composite			Common K, A K=0.4 A=0.4
	Optimal K	Relative Efficiency	Optimal K	A	Relative Efficiency	Relative Efficiency
In Labour Force	0.9	146.6	0.9	0.1	147.9	113.3
Employed	0.9	151.0	0.9	0.1	152.3	114.1
Employed Agriculture	0.9	234.7	0.9	0.0	234.7	112.3
Employed Non-Agriculture	0.9	154.0	0.9	0.1	155.2	114.1
Unemployed	0.4	106.0	0.6	0.2	106.4	102.9

Relative efficiency is with respect to the simple ratio estimator and is defined as 100 times the ratio of appropriate variances.

APPENDIX I

Derivation of Bias of the Composite Estimator:

As defined in (2.2), the AK composite estimator of Y_m is given by:

$$y'_m = (1 - K + A)y_{m,1}/6 + (1 - K - A/5) \sum_{j=2}^{6} y_{m,j}/6 + K(y'_{m-1} + d_{m,m-1}). \quad (A1.1)$$

It may be noted that the simple ratio estimator now employed in LFS is the average of the six panel estimators and is given by:

$$\bar{y}_m = \sum_{i=1}^{6} y_{m,i}/6. \quad (A1.2)$$

From (2.3), the bias of the i-th panel estimator equals α_i so that:

$E(y_{m,i}) = Y_m + \alpha_i$, recalling that the bias is independent of m. Hence,

$$E(\bar{y}_m) = Y_m + \sum_{i=1}^{6} \alpha_i/6 = Y_m + \bar{\alpha} \text{ (say)}.$$

In later developments we assume that $\bar{\alpha} = 0$.

The composite estimator may be rewritten as:

$$y'_m = y_m + K(y'_{m-1} + d_{m,m-1}), \tag{A1.3}$$

where

$$y_m = (1 - K + A)y_{m,1}/6 + (1 - K - A/5) \sum_{j=2}^{6} y_{m,j}/6$$

$$= (1 - K)\bar{y}_m + A(y_{m,1} - \bar{y}_m)/5.$$

Therefore

$$E(y_m) = (1 - K)(Y_m + \bar{\alpha}) + (A/5)(\alpha_1 - \bar{\alpha}). \tag{A1.4}$$

When $\bar{\alpha} = 0$, it simplifies to

$$E(y_m) = (1 - K)Y_m + (A/5)\alpha_1.$$

Using the definition of $d_{m,m-1}$ given in (2.1), we have

$$E(d_{m,m-1}) = E[\sum_{j=2}^{6} (y_{m,j} - y_{m-1,j-1})/5]$$

$$= (Y_m - Y_{m-1}) + (\alpha_6 - \alpha_1)/5. \tag{A1.5}$$

Now y'_m may be expanded by applying (A1.3) recursively and it is found that, up to n months back:

$$y'_m = y_m + Ky_{m-1} + K^2y_{m-2} + \dots + K^{n-1}y_{m-n+1} + K^ny'_{m-n}$$

$$+ Kd_{m,m-1} + K^2d_{m-1,m-2} + \dots\dots + K^nd_{m-n+1,m-n}. \tag{A1.6}$$

The expected value of y'_m may be readily obtained from (A1.4) and (A1.6) as below:

$$E(y'_m) = (1 - K)\ [Y_m + KY_{m-1} + K^2Y_{m-2} + \dots + K^{n-1}Y_{m-n+1}] + K^nE(y'_{m-n})$$

$$+ [(1 - K)\ \bar{\alpha} + (A/5)(\alpha_1 - \bar{\alpha})](1 - K^n)/(1 - K)$$

$$+ K(Y_m - Y_{m-1}) + K^2(Y_{m-1} - Y_{m-2}) + \dots + K^n(Y_{m-n+1} - Y_{m-n})$$

$$+ [(\alpha_6 - \alpha_1)/5]K(1 - K^n)/(1 - K)$$

$$= Y_m + K^n[E(y'_{m-n}) - Y_{m-n})]$$

$$+ [(1 - K)\bar{\alpha} + (A/5)(\alpha_1 - \bar{\alpha}) + K(\alpha_6 - \alpha_1)/5](1 - K^n)/(1 - K)$$

$$= Y_m + K^n[E(y'_{m-n}) - Y_{m-n}]$$

$$+ [(1 - K - A/5)\bar{\alpha} + (A/5)\alpha_1 + K(\alpha_6 - \alpha_1)/5](1 - K^n)/(1 - K) \qquad (A1.7)$$

which simplifies for sufficiently large n and for the case $\bar{\alpha} = 0$ to

$$E(y'_m) = Y_m + [A\alpha_1 + K(\alpha_6 - \alpha_1)]/[5(1 - K)]. \qquad (A1.8)$$

Since the bias of y'_m under the model assumed in this paper is independent of m, the difference between composite estimates r months apart is unbiased, i.e.,

$$E(y'_m - y'_{m-r}) = Y_m - Y_{m-r} \text{ for all r.} \qquad (A1.9)$$

APPENDIX II

Derivation of the Variance of the Composite Estimator

We assume that the composite estimators (see (2.2)) have become sufficiently stable over time and hence we shall assume that $V(y'_{m-1}) = V(y'_m)$. Since all correlations 12 or more months apart are assumed to be zero, we shall assume that the LFS composite estimators have become stable after 12 months.

Taking the variance of both sides of (A1.3) and applying the above assumption; one may solve for $V(y'_m)$ to find that:

$$V(y'_m) = [V(y_m) + K^2V(d_{m,m-1}) + 2KCov(y_m,\ d_{m,m-1})$$

$$+ 2KCov(y_m,\ y'_{m-1}) + 2K^2Cov(d_{m,m-1},\ y'_{m-1})]/(1 - K^2). \qquad (A2.1)$$

To eliminate y'_{m-1} on the right side of (A2.1), we apply (A1.6), replacing m by (m - 1) and n by 12 to obtain:

$$y'_{m-1} = \sum_{g=1}^{12} (K^{g-1}y_{m-g} + K^g d_{m-g,m-g-1}) + K^{12}y'_{m-13}. \qquad (A2.2)$$

Substituting (A2.2) in (A2.1) and dropping zero terms, we have

$$V(y'_m) = \{V(y_m) + K^2V(d_{m,m-1}) + 2KCov(y_m,\ d_{m,m-1})$$

$$+ 2 \sum_{g=1}^{12} K^g[Cov(y_m,\ y_{m-g}) + KCov(d_{m,m-1},\ y_{m-g})$$

$$+ KCov(y_m,\ d_{m-g,m-g-1})$$

$$+ K^2Cov(d_{m,m-1},\ d_{m-g,m-g-1})]\}/(1 - K^2). \qquad (A2.3)$$

We give the expressions for the variances and covariances on the right side of (A2.3), which may be readily derived by considering (2.1) and (A1.3).

$$V(y_m) = [(1 - K)^2/6 + A^2/30]\sigma^2, \qquad (A2.4)$$

which simplifies to $\sigma^2/6$ when $A = K = 0$; i.e.,

$$V(\bar{y}_m) = \sigma^2/6, \qquad (A2.4a)$$

the variance of the current LFS estimator.

$$V(d_{m,m-1}) = 2\sigma^2(1 - \rho_1)/5. \tag{A2.5}$$

$$Cov(y_m, d_{m,m-1}) = (1 - K)(1 - \rho_1)\sigma^2/6 - A(1 - \rho_1)\sigma^2/30. \tag{A2.6}$$

To derive the remaining covariances in (A2.3), which involve 'g', an indicator function I(a, b) shall be defined by:

$$\begin{aligned} I(a, b) &= 1 \text{ if } a \leq b \\ &= 0 \text{ otherwise.} \end{aligned}$$

By considering the definitions of y_m in (A1.3), $d_{m,m-1}$ in (2.1) and the corresponding expressions for month (m - g), one would find that the following covariances would be required to derive the remaining covariance of (A2.3).

$$\begin{aligned}
Cov(\bar{y}_m, \bar{y}_{m-g}) &= (\sigma^2/36)[(6 - g)\rho_g I(g, 5) + (6 - |\gamma - 6|)\gamma_g], \\
Cov(\bar{y}_m, y_{m-g,1}) &= (\sigma^2/6)[\rho_g I(g, 5) + \gamma_g I(6, g)], \\
Cov(y_{m,1}, \bar{y}_{m-g}) &= (\sigma^2/6)\gamma_g I(g, 6), \\
Cov(y_{m,1}, y_{m-g,1}) &= \sigma^2 I(g, 6)I(6, g)\gamma_6 \text{ (= 0 except when } g = 6), \\
Cov(y_{m-1,6}, \bar{y}_{m-g}) &= (\sigma^2/6)[\rho_{g-1} I(g, 6) + \gamma_{g-1} I(7, g)], \\
Cov(y_{m-1,6}, y_{m-g,1}) &= \sigma^2[\rho_{g-1} I(g, 6)I(6, g)], \\
Cov(\bar{y}_m, y_{m-g-1,6}) &= (\sigma^2/6)\gamma_{g+1} I(g, 5), \\
Cov(y_{m,1}, y_{m-g-1,6}) &= \sigma^2\gamma_1 I(g, 0)I(0, g) \text{ (= 0 for } g \geq 1).
\end{aligned} \tag{A2.7}$$

The four covariances of (A2.3) that involve g may be readily defined and are found to be as follows:

$$Cov(y_m, y_{m-g}) = \sigma^2\rho_g I(g, 5)[(1 - K)^2(6 - g)/36$$

$$+ A(1 - K)(g - 3)/90 - gA^2/900]$$

$$+ \sigma^2\gamma_g[(1 - K)^2(6 - |g - 6|)/36$$

$$+ A(1 - K)(|g - 6| - 3)/90 - |g - 6| A^2/900]$$

$$+ \sigma^2\gamma_g I(g, 6)I(6, g)A(1 - K + A)/30, \qquad (A2.8)$$

$$\mathrm{Cov}(d_{m,m-1}, y_{m-g}) = \sigma^2(\rho_g - \rho_{g-1})I(g, 5)[(1 - K)(6 - g)/30 + gA/150]$$

$$+ \sigma^2(\gamma_g - \gamma_{g-1})[(1 - K)(6 - |g - 6|)/30 + |g - 6|A/150$$

$$- (1 - K + A)I(g, 6)/30], \qquad (A2.9)$$

$$\mathrm{Cov}(y_m, d_{m-g,m-g-1}) = \sigma^2(\rho_g - \rho_{g+1})I(g, 5)(1 - K - A/5)(5 - g)/30$$

$$+ \sigma^2(\gamma_g - \gamma_{g+1})[(1 - K - A/5)(6 - I(6, g)$$

$$- |g - 6|)/30 + AI(g, 5)/25], \qquad (A2.10)$$

$$\mathrm{Cov}(d_{m,m-1}, d_{m-g,m-g-1}) = \sigma^2[(5 - g)(2\rho_g - \rho_{g-1} - \rho_{g+1})I(g, 5)$$

$$+ (5 - |g - 6|)(2\gamma_g - \gamma_{g-1} - \gamma_{g+1})]/25. \qquad (A2.11)$$

Hence, $V(y'_m)$ can be expressed as $aA^2 + bA + c = f(A)$ where a, b and c are functions of K, ρ's and γ's. It can be shown that $a \geq 0$. The values of A that minimize the variance of AK estimator was determined for K = 0(0.1)0.9. Among these (A, K)'s, the optimal value of (A, K) was selected and is presented in Table 4A.

APPENDIX III

Derivation of the variance of $Y'_m - Y'_{m-1}$

From (A1.3)

$$y'_m = y_m + K(y'_{m-1} + d_{m,m-1}), \text{ or}$$

$$y'_m - Ky'_{m-1} = y_m + Kd_{m,m-1},$$

whence

$$(1 + K^2)V(y'_m) - 2KCov(y'_m, y'_{m-1}) = V(y_m) + 2KCov(y_m, d_{m,m-1}) + K^2V(d_{m,m-1}).$$

When $K \neq 0$, $Cov(y'_m, y'_{m-1})$ may be obtained from the above and upon substitution of (A2.4), (A2.6) and (A2.5), and from the fact that $V(y'_m - y'_{m-1}) = 2V(y'_m) - 2Cov(y'_m, y'_{m-1})$, one may find that for $K \neq 0$:

$$V(y'_m - y'_{m-1}) = \sigma^2[A^2/30 - (1 - \rho_1)KA/15 + (1 - K)^2/6 + (1 - \rho_1)K(K + 5)/15]/K - (1 - K)^2V(y'_m)/K. \quad \text{(A3.1)}$$

When $K = 0$,

$$y'_m = (1 - A/5)\bar{y}_m + Ay_{m,1}/5,$$

$$Cov(y'_m, y'_{m-1}) = Cov[(1 - A/5)\bar{y}_m + Ay_{m,1}/5, (1 - A/5)\bar{y}_{m-1} + Ay_{m-1,1}/5]. \quad \text{(A3.2)}$$

Thus for $K = 0$, we have:

$$V(y'_m - y'_{m-1}) = \sigma^2[(1/15 + \rho_1/450 + \gamma_1/90)A^2 + 2(\rho_1 - \gamma_1)A/45 + 1/3 - (5\rho_1 + \gamma_1)/18]. \quad \text{(A3.3)}$$

REFERENCES

[1] Bailar, B.A. (1975), "The Effects of Rotation Group Bias on Estimates from Panel Surveys", Journal of the American Statistical Association,

70, 23-29.

[2] Ghangurde, P.D. (1982), "Rotation Group Bias in the LFS Estimates", Survey Methodology, 8, 86-101.

[3] Gurney, M., and Daly, J.F. (1965), "A Multivariate Approach to Estimation in Periodic Sample Surveys", Proceedings of the Social Statistics Section, American Statistical Association, 242-257.

[4] Huang, E., and Ernst, L. (1981), "Comparison of an Alternate Estimator to the Current Composite Estimator in CPS", Proceedings of the Section on Survey Research Methods, American Statistical Association, 303-308.

[5] Lee, H. (1983), "Estimation of Panel Correlations in the Canadian Labour Force Survey", Technical Memorandum (in Preparation), Census and Household Survey Methods Division, Statistics Canada.

[6] Tessier, R. and Tremblay, V. (1976), "Findings on Rotation Group Biases", Technical Memorandum, Statistics Canada.

[7] Statistics Canada (1976), Methodology for the Canadian Labour Force Survey, Catalogue 71-526, Occasional.