EVALUATION OF COMPOSITE ESTIMATION FOR THE CANADIAN LABOUR FORCE SURVEY

S. Kumar and H. Lee, Statistics Canada

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 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 for estimates of level and estimates of 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 (henceforth referred to as simple estimator). The study is based on the province of Ontario data for 1980 - 1981.

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 estimator of Y_m based on the i^{th} panel (i = 1,2,..,6). Here i^{th} pannel 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.

$d_{m,m-1}$ = estimate 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)

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

y'_m = AK composite estimator of Y_m defined as

(2.1)
$$y'_m = \frac{1}{6}(1 - K + A)y_{m,1} + \frac{1}{6}\left(1 - K - \frac{A}{5}\right)\sum_{j=2}^{6} y_{m,j} + K(y'_{m-1} + d_{m,m-1})$$

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

The equation (2.1) defines a class of estimators referred to as AK composite estimators. The estimators obtained by taking A = 0 in (2.1) are referred to as K composite estimators. The simple estimator, to be denoted by y_m, can ne obtained by taking A = 0 and K = 0 in (2.1). We investigate the relative performance of the optimal (minimum variance or minimum mean square error) AK composite, K composite and simple estimators.

We assume that 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,

(2.2)
$$\alpha_i = E(y_{m,i}) - Y_m$$

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

3. Assumptions

The expression for the variance of y'_m, $V(y'_m)$, involves the variances and covariances of various panel estimators (see appendix I). We assume the following variance - covariance structure for various panel estimators. These assumptions conform to the LFS rotation pattern.

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

(ii) $Cov(y_{m,i},\, y_{m-j,\, i-j}) = \rho_j\sigma^2$,
where j = 1, 2, ..5 and $j < i \le 6$.

Here $\rho_j\sigma^2$ is the covariance between overlapping rotation groups j months apart and is assumed to be stationary, i.e., it is a function of j and not of m.

(iii) $Cov(y_{m,i},\, y_{m-j,\, 6+i-j}) = \gamma_j\sigma^2$

where (a) $1 \le i \le j \le 6$, or
(b) $7 \le j \le 11$, and $j - 5 \le i \le 6$

That is, $\gamma_j\sigma^2$ is the covariance between estimators based on a rotation group and its immediate predecessor rotation group j months apart and is assumed to be stationary as before. It is reasonable to assume that γ's will be positive since generally households on their final exit from the sample are replaced by their neighbours.

(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 $i \ne j$, $Cov(y_{m,i},\, y_{m-1,j})$ for i = 1, $j \ne 6$, and $i \ne 1$, $j \ne i - 1$, and $Cov(y_{m,i},\, y_{m-g,\, j})$ for $g \ge 12$.

These and all other covariances not defined above 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 Appendix I.

4. Results and Discussion

The quantities ρ_j and γ_j in the expression 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, under the assumptions (see section 3), $\rho_j = 0$ for $j \geq 6$ and $\gamma_j = 0$ for $j \geq 12$. Estimates of ρ_j, $\hat{\rho}_j$, are given in table 1. 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 $\hat{\rho}_j$ $(j = 1,2,..,5)$ is a decreasing function of j for all the five characteristics. This is consistent with what we expect intutively about the behaviour of ρ_j's. Also ρ_j's are high for all the characteristics except unemployed.

Table 1 also gives the estimates $\hat{\gamma}_j$. The estimates $\hat{\gamma}_5$ and $\hat{\gamma}_{11}$ were obtained respectively by interpolating and extrapolating other $\hat{\gamma}_j$'s. Intuitevely, 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.

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

Table 2 gives 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 (2.1), (ii) optimal K composite estimator (obtained by taking A = 0 in (2.1) and having minimum variance among all estimators in this subclass), and (iii) the simple 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 (2.1). Similar definition applies to the term 'optimal K'. Table 2 (computed using $\hat{\gamma}_j$'s given in table 1) 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 also given in table 2. Note that the optimal K's and optimal (K,A)'s in the two cases are different. Comparison of the corresponding relative efficiencies in these two cases shows that positive γ's have negative effect on the reduction in variance, i.e., 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 estimates of level is their bias as compared to the simple estimator. Thus comparing the variances of biased estimators can sometimes result in erroneous conclusions about the relative performance of these estimates. It is appropriate to examine the mean square error in the case of biased estimates. 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 estimator 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 3. 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 α_i's. For each of the remaining two characteristics, $\hat{\alpha}_i$'s do not deviate too much.

Tables 4 and 5 give the optimal K, the optimal (K,A) and results of comparing mean square errors. Two criteria of optimality are used. One is based on the concept of minimum variance (as is the case for table 2), and the other is based on the concept of minimizing the 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 tables 4&5 is computed by using $\hat{\alpha}_1$ and $\hat{\alpha}_6$ (given in table 3) instead of α_1 and α_6 in the above formula. Now we discuss the results of tables 4 and 5.

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

The K composite estimates (considered in table 2 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 that the simple estimator, i.e., the gain in relative efficiency is insignificant. The difference in the corresponding relative efficiency results in tables 2 and 4 is due to the different relative efficiency definitions used for the two tables. For table 2, relative efficiency is defined as the ratio of appropriate variances whereas for tables 4 and 5, 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.

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", and "employed non-agriculture". Also, there is large bias in these cases. The gain in relative efficiency for the characteristic "employed agriculture" is moderate whereas the corresponding gain for the characteristic "unemployed" is nominal. The slight difference in the optimal (K,A) values (corresponding to the minimum variance) in tables 2 and 5 is due to the minor differences in the computational techniques used for generating these tables.

The results in tables 4&5 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 5.

We note, from the expression for $E(y'_m)$ given earlier, that $y'_m - y'_{m-1}$ is an unbiased estimator of $Y_m - Y_{m-1}$, i.e., K or AK composite estimates of change are unbiased. Table 6 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 values of K or (K,A) are characteristic dependent. Thus the additive property of the estimates is not preserved. 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 5 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 6 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".

TABLE 1

Estimated Correlations[1] ρ's and γ's (1980-1981 Ontario)

Characteristics	i	1	2	3	4	5	6	7	8	9	10	11
In Labour Force	$\hat{\rho}_i$	.843	.782	.717	.674	.631	-	-	-	-	-	-
	$\hat{\gamma}_i$	.161	.141	.128	.133	.135	.136	.125	.127	.124	.122	.127
Employed	$\hat{\rho}_i$	.852	.779	.709	.664	.619	-	-	-	-	-	-
	$\hat{\gamma}_i$	.164	.136	.142	.142	.146	.149	.148	.150	.153	.141	.148
Employed - Agriculture	$\hat{\rho}_i$	.955	.926	.901	.861	.821	-	-	-	-	-	-
	$\hat{\gamma}_i$	.477	.483	.474	.486	.480	.474	.459	.429	.394	.323	.252
Employed Non-Agriculture	$\hat{\rho}_i$	.861	.791	.724	.678	.632	-	-	-	-	-	-
	$\hat{\gamma}_i$	.184	.150	.147	.157	.162	.167	.166	.169	.174	.156	.166
Unemployed	$\hat{\rho}_i$	.580	.445	.334	.286	.238	-	-	-	-	-	-
	$\hat{\gamma}_i$	.141	.074	.076	.063	.057	.051	.045	.060	.077	.136	.074

[1] ρ_i (i = 6,7, ...,11) is equal to zero (see section 4).

TABLE 2

The Optimal (K,A) and K, and the Relative

Efficiencies[2] of K Composite and AK Composite Estimators

Characteristics	Using $\hat{\gamma}_i$ values given in Table 1					$\hat{\gamma}_i$=0 for all i				
	K Composite		AK Composite			K Composite		AK Composite		
	Optimal K	Relative Efficiency	Optimal K	A	Relative Efficiency	Optimal K	Relative Efficiency	Optimal K	A	Relative Efficiency
In Labour Force	0.7	118.8	0.8	0.48	128.4	0.7	125.5	0.8	0.50	138.4
Employed	0.7	118.5	0.8	0.49	128.1	0.7	125.3	0.8	0.51	137.9
Employed Agriculture	0.8	120.6	0.8	0.38	126.9	0.8	167.3	0.9	0.46	187.9
Employed Non-Agriculture	0.7	119.4	0.8	0.47	129.3	0.7	126.9	0.8	0.49	140.2
Unemployed	0.3	102.8	0.5	0.38	105.2	0.4	104.4	0.6	0.51	108.4

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

TABLE 3

Estimates (in thousands) of Rotation Group Bias α_i[3]

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

[3] 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 4

Comparison of Simple and K Composite Estimators [4,5]

Characteristics	Simple Monthly level		K Composite									
			Minimum Mean Square Error					Minimum Variance				
	Est.	Var.	K	Var.	Bias	Mse.	R.E.	K	Var.	Bias	Mse	R.E.
In Labour Force	4480.7	432.0	0.0	432.0	0.0	432.0	100.00	0.7	363.8	66.8	4284.5	9.0
Employed	4186.0	473.3	0.0	473.3	0.0	473.3	100.0	0.7	399.6	73.0	5732.2	8.3
Emp. Agr.	142.0	85.7	0.6	75.6	1.4	77.6	110.5	0.8	71.1	3.7	85.1	100.7
Emp. Non-Agr.	4043.9	498.9	0.0	498.9	0.0	498.9	100.0	0.7	417.8	70.8	5436.1	9.2
Unemployed	294.8	117.5	0.2	114.9	-0.7	115.4	101.9	0.3	114.3	-1.1	115.7	101.6

[4] Estimates are in thousands, var. and Mse. are in millions.

[5] Relative Efficiency (in tables 4 and 5) is with respect to the simple estimator and is defined as 100 times the ratio Mse (simple estimator)/Mse (K or AK Composite).

TABLE 5
Comparison of Simple and AK Composite Estimators

Characteristics	AK Composite															
	Minimum Mean Square Error						Minimum Variance						K=A=.4 for all characteristics			
	K	A	Var.	Bias	Mse	R.E.	K	A	Var.	Bias	Mse.	R.E.	Var.	Bias	Mse.	R.E.
In Labour Force	.7	.7	358.1	3.7	371.5	116.3	.8	.5	336.5	46.9	2532.4	17.1	391.8	1.0	392.9	110.0
Employed	.8	.9	397.7	-2.4	403.2	117.4	.8	.5	369.5	54.3	3320.9	14.3	428.9	2.0	432.8	109.4
Emp. Agr.	.8	.6	68.7	1.2	70.2	122.2	.8	.4	67.6	2.1	71.8	119.4	80.8	0.1	80.8	106.1
Emp. Non-Agr.	.8	.9	418.0	-2.3	423.3	117.9	.8	.5	385.9	52.7	3161.7	15.8	452.8	1.9	456.4	109.5
Unemployed	.4	.4	112.5	-0.9	113.3	103.7	.5	.4	111.7	-1.6	114.4	102.7	112.5	-0.9	113.3	103.7

TABLE 6
Relative Efficiency[6] 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	Optimal 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
Emp. Agr.	0.9	234.7	0.9	0.0	234.7	112.3
Emp. Non-Agr.	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

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

APPENDIX I

1. Expression for the Variance of y'_m

We assume that the composite estimators have become sufficiently stable over time and hence we can take $V(y'_m) = V(y'_{m-1})$. We express (2.1) as

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

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

then

(4.3) $V(y'_m) = [V(y_m) + K^2 V(d_{m,m-1}) + 2K\,Cov(y_m, d_{m,m-1}) + 2K\,Cov(y_m, y'_{m-1}) + 2K^2\,Cov(d_{m,m-1}, y'_{m-1})]/(1-K^2)$

To eliminate y' on the right side of (4.3), we expand as follows:

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

Substituting (4.4) in (4.2) and dropping zero terms, we have

(4.5) $V(y'_m) = [V(y_m) + K^2 V(d_{m,m-1}) + 2K\,Cov(y_m, d_{m,m-1}) + 2 \sum_{g=1}^{12} K^g \{Cov(y_m, y_{m-g}) + K\,Cov(d_{m,m-1}, y_{m-g}) + K\,Cov(y_m, d_{m-g,m-g-1}) + K^2\,Cov(d_{m,m-1}, d_{m-g,m-g-1})\}]/(1-K^2)$

We give the expressions for the variances and covariances on the right side of (4.5).

(4.6) $V(y_m) = [5(1-K)^2 + A^2]\,\sigma^2/30$

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

(4.8) $Cov(y_m, d_{m,m-1}) = [5(1-K)(1-\rho_1)\sigma^2 - A(1-\rho_1)\sigma^2]/30$

We define the indicator function I(a,b) as follows:

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

Let $g' = g-6$ and $\rho_0 = 1$. Also, note that $\rho_g = 0$ for $g \geq 6$ and $\gamma_g = 0$ for $g \geq 12$. Then

(4.9) $Cov(y_m, y_{m-g}) = \{25(1-K)^2(6-g) + 10A(1-K)(g-3) - gA^2\}\,\rho_g I(g,5)\sigma^2/900 + \{25(6-|g'|)(1-K)^2 + 10(|g'|-3)A(1-K) - |g'|A^2\}\gamma_g\,\sigma^2\{1-I(g,6)I(6,g)\}/900 + \{5(1-K)^2 + A^2\}\gamma_6\sigma^2\,I(g,6)I(6,g)/30.$

(4.10) $\mathrm{Cov}(d_{m,m-1}, y_{m-g}) = \{5(6-g)(1-K) + gA\}\ I(g,5)(\rho_g - \rho_{g-1})\rho^2/150 + [\{5(g-1)(1-K) - (g-1)A\}\ I(g,6) + \{5(12-g)(1-K) + (g-6)A\}\ I(7,g)]\ (\gamma_g - \gamma_{g-1})\sigma^2/150$

(4.11) $\mathrm{Cov}(y_m, d_{m-g,m-g-1}) = \{5(5-g)(1-K) - (5-g)A\}\ I(g,5)(\rho_g - \rho_{g+1})\sigma^2/150 + [\{5g(1-K) - A(g-6)\}\ I(g,5) + \{5(11-g)(1-K) - (11-g)A\}\ I(6,g)]\ (\gamma_g - \gamma_{g+1})\sigma^2/150$

(4.12) $\mathrm{Cov}(d_{m,m-1}, d_{m-g,m-g-1}) = \sigma^2\ [(5-g)(2\rho_g - \rho_{g-1} - \rho_{g+1})\ I(g,5) + (5-|g-6|)\ (2\gamma_g - \gamma_{g-1} - \gamma_{g+1})]/25.$

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 (0.1)0.9. Among these (A,K)'s, the optimal value of (A,K) was selected and is presented in table 2.

2. Expression for the bias of y'_m

Let α_i be the bias of the estimator $y_{m,i}$. Note that the bias is assumed to be independent of m and is a function of the rotation group. Formally,

$E(y_{m,i}) = Y_m + \alpha_i$. Then

$E(y_m) = (1-K)Y_m + (1-K) \sum_{i=1}^{6} \alpha_i/6 + A(\alpha_1 - \sum_{i=2}^{6} \alpha_i/5)/6 = (1-K)Y_m + \alpha(\text{say}).$

$E(d_{m,m-1}) = Y_m - Y_{m-1} + (\alpha_6 - \alpha_1)/5.$ Hence

$E(y'_m) = Y_m + \{\alpha + K(\alpha_6 - \alpha_1)/5\}\ (1+K+\ldots+K^{n-1}) + K^n E(y'_{m-n} - Y_{m-n}).$

Hence, for sufficiently large n

$E(y'_m) = Y_m + \sum_{i=1}^{6} \alpha_i/6 + A\ (\alpha_1 - \sum_{i=2}^{6} \alpha_i/5)/(6(1-K)) + K(\alpha_6 - \alpha_1)/(5(1-K)).$

In our study, it is assumed that $\sum_{i=1}^{6} \alpha_i = 0$.

For this case, we have

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

3. Expression for the variance of $y'_m - y'_{m-1}$

For K=0,

$V(y'_m - y'_{m-1}) = \sigma^2\ \{(30 + \rho_1 + 5\gamma_1)A^2 + 20(\rho_1 - \gamma_1)A + 150 - 25\ (5\rho_1 + \gamma_1)\}/450,$

for $K \neq 0$,

$V(y'_m - y'_{m-1}) = \sigma^2\ \{A^2 - 2(1-\rho_1)KA + 5(1-K)^2 + 2(1-\rho_1)K(K+5)\}/30K - (1-K)^2\ V(y'_m)/K.$

Acknowledgement The authors wish to thank D. Drew, G.B. Gray and M.P. Singh, Statistics Canada for helpful discussions during the progress of this work.

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", American Statistical Association: Proceedings of the Social Statistics Section of the American Statistical Association, 242-257.
4. Huang, E., and Ernst, L. (1981), "Comparison of an Alternative Estimator to the Current Composite Estimator in CPS", Proceedings of the Section on Survey Research Methods of the 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", Internal Technical Memorandum, Statistics Canada.
7. Statistics Canada (1976), Methodology for the Canadian Labour Force Survey, catalogue 71-526, occasional.