SSC Annual Meeting, June 1996
Proceedings of the Survey Methods Section

COMPOSITE ESTIMATION IN THE PRESENCE OF ROTATION GROUP BIAS

A.C. Singh, J.G. Gambino, and E.J. Chen[1]

ABSTRACT

For rotating panel surveys, the usual composite estimator known as the *AK*-composite estimator, based on panel estimates for current and previous time points, requires the assumption that rotation group bias (*RGB*) is either absent or negligible. In this article, we propose an alternative estimator, termed *ABK*-composite, which does not require *RGB* to be negligible. We consider both univariate and multivariate composite estimation. It is shown that there is a trade-off between bias and variance in moving from *AK* to *ABK*. In the presence of *RGB*, the *ABK*-estimator continues to be unbiased but with larger variance than the *AK*-estimator which becomes biased but with smaller variance. However, if *RGB* is negligible, then *ABK* is a less efficient alternative to *AK*. Through numerical examples, it is shown that there may be substantial loss in efficiency with *ABK* compared to *AK*. It is recommended that measures such as improved interviewer training, and panel-level nonresponse adjustment be taken to make panels more balanced, so that *RGB* could be deemed negligible.

KEY WORDS: *AK*-Composite; *ABK*-Composite; Panel Balancing; Rotating Panel Surveys.

RÉSUMÉ

Pour les enquêtes où il y a renouvellement du panel, l'estimateur combiné habituel, appelé estimateur combiné *AK*, qui repose sur l'estimation du panel au temps présent et au temps antérieur, suppose un biais de renouvellement (*BR*) inexistant ou négligeable. Les auteurs proposent un nouvel estimateur, baptisé estimateur combiné *ABK*, en vertu duquel le *BR* n'a pas besoin d'être négligeable. Ils examinent l'estimation combinée à une et à plusieurs variables. On constate qu'il faut faire un compromis entre le biais et la variance lorsqu'on passe de l'estimateur *AK* à l'estimateur *ABK*. En présence de *BR*, l'estimateur *ABK* ne donne aucun biais mais sa variance est plus importante alors que l'estimateur *AK* est biaisé mais présente une plus faible variance. Lorsque le *BR* est négligeable cependant, l'estimateur *ABK* se montre moins efficace que l'estimateur *AK*. Des exemples numériques montrent que l'estimateur *ABK* pourrait entraîner une perte d'efficacité appréciable par rapport à l'estimateur *AK*. On recommande l'adoption de certaines mesures comme une meilleure formation des intervieweurs et un ajustement en fonction des non-réponses au niveau des panels afin de mieux équilibrer les panels pour que le *BR* reste négligeable.

MOTS CLÉS: Estimateur combiné *AK*; estimateur combiné *ABK*; équilibrage des panels; enquête avec renouvellement du panel.

1. INTRODUCTION

With rotating panel surveys, the usual estimates of level and change based only on cross-sectional data can be made more efficient by incorporating in the estimates correlated information from other time points because of overlapping samples. The traditional estimator is the *AK*-estimator of Gurney and Daly (1965) which can be developed from elementary (or panel level) estimates under a general linear model framework; see Smith (1978) and Wolter (1979). Let t, t' denote the current and previous time points and let y, m denote respectively the study variable at t and the matched (backward with the t'sample) subsample of the t'-sample. Similarly, y', m' are defined where m', for example, now denotes the matched (forward with the t-sample) subsample of the t'-sample. The AK-composite estimator, $\hat{\theta}^{AK}_{yt}$ of the total θ_{yt} for the study variable y at time t is given by

$$\hat{\theta}^{AK}_{yt} = \hat{\theta}^{GR}_{yt} + K\left[\hat{\theta}^{AK}_{y't'} + (\hat{\theta}^{GR}_{ytm} - \hat{\theta}^{GR}_{y't'm'}) - \hat{\theta}^{GR}_{yt}\right] + A(\hat{\theta}^{GR}_{yt} - \hat{\theta}^{GR}_{ytm}), \quad (1.1a)$$

$$= \hat{\theta}^{GR}_{yt} + K(\hat{\theta}^{AK}_{y't'} - \hat{\theta}^{GR}_{y't'm'}) + (A-K)(\hat{\theta}^{GR}_{yt} - \hat{\theta}^{GR}_{ytm}), \quad (1.1b)$$

[1] A.C. Singh, Methodology Research Advisory Group; J.G. Gambino, Survey and Analysis Methods Development Section, and E.J. Chen, Social and Economic Studies Division, Statistics Canada, Ottawa, K1A 0T6.

where, $\hat{\theta}^{GR}_{ytm}$, for example, denotes the usual generalized regression (*GR*) estimator based on the cross-sectional subsample *m*. In the sequel, it will be convenient at times to use the following simpler (but less rigorous) notation than that of 1.1(a, b). Thus we will also denote the full sample estimates corresponding to variables *y* and y' by F_y and F'_y respectively, and the partial sample estimates by $P_{y(m)}$ and $P'_{y(m)}$ respectively, where the subscript signifies the matched subsample. Note that all the above estimates are *GR* except for F'_y which will be *AK*. Now, the equation (1.1) can be re-expressed as

$$AK(y) = F_y + K[F'_y + (P_{y(m)} - P'_{y(m)}) - F_y] + A(F_y - P_{y(m)}), \quad (1.2a)$$

$$= F_y + K(F'_y - P'_{y(m)}) + (A-K)(F_y - P_{y(m)}). \quad (1.2b)$$

The *AK*-estimator is an optimal regression estimator in a linear class in that the coefficients *A* and *K* are chosen to minimize the variance. It is recursive in nature. For any given time *t*, it is not fully optimal in the sense of combining optimally all the correlated information from the past. However, it does combine optimally information from *t* and t'; the estimator, $\hat{\theta}^{AK}_{y't'}$ at t' represents a condensed form of all the past information. Computationally, the *AK*-estimator is a two-step estimator where in the first step, a *GR*-estimator of θ_{yt} is computed using the usual time *t*-predictor zero functions $(\theta_{xt} - \hat{\theta}^{HT}_{xt})$ where *x* may denote a set of demographic characteristics such as age-sex groups, and in the second step, optimal regression is used to combine $\hat{\theta}^{GR}_{yt}$ with two additional predictor zero functions $(\hat{\theta}^{AK}_{y't'} - \hat{\theta}^{GR}_{y't'm'})$ and $(\hat{\theta}^{GR}_{yt} - \hat{\theta}^{GR}_{ytm})$. The *K*-composite estimator, a predecessor of *AK*, is a special case of *AK* obtained by setting *A*=0. Thus, in the *K*-composite estimator of Hansen *et al.* (1953), the additional predictor zero function is simply the difference of the two additional predictors used in *AK*. A further development of *AK* which can be termed the *AK*-calibration estimator, was suggested by Fuller (1990), in which a set of composite weights are produced for estimation of all study variables. For this purpose, for a few linearly independent key variables, the *AK*- estimators are first obtained, and then these estimates are used as additional auxiliary population totals in the *GR*-method to find final calibrated weights.

Often several characteristics are measured on the same sample unit. It is then possible to take advantage of correlation both over time and across characteristics. Thus, additional predictions $(\hat{\theta}^{AK}_{z't'} - \hat{\theta}^{GR}_{z't'm'})$ and $(\hat{\theta}^{GR}_{zt} - \hat{\theta}^{GR}_{ztm})$ corresponding to a set of key study variables (*y*, *z*, etc.) can be used together. This will make the *AK*-estimator multivariate in nature. Note that the traditional *AK* is, however, univariate in nature.

All the above methods assume that the predictors are zero in expectation. This will be so if the rotation group bias (*RGB*) is zero. The bias due to rotation group (or panel) difference as defined by Bailar (1975) implies that individual panel means depend on time-in-sample, and, therefore, vary from panel to panel. Moreover, it is generally the birth panel that differs most from others. Now it is known from the empirical studies of Huang and Ernst (1981) and Kumar and Lee (1983) that the magnitude of the estimated *RGB* for the optimal *AK*-composite estimator is found to be less than that for the optimal *K*-composite estimator. This can be expected in view of the conditional unbiasedness of optimal regression estimators (Valliant, 1993) which gives rise to sample balance with respect to predictors. In particular, for *AK*, bias in the direction of the additional predictor $(F_y - P_{y(m)})$ reflecting the difference between birth and nonbirth panels is eliminated. Thus, the *AK*-estimator improves the *K*-estimator both with respect to bias and variance. However, there still remains some *RGB* in *AK* which may be significant.

The purpose of this article is to propose a bias-adjusted *AK* estimator (denoted by *ABK*) that is free from *RGB* under certain commonly made assumptions. These assumptions state that *RGB* is additive, constant over time, and vanishes when averaged over panels. The proposed *ABK*-estimator is derived from first principles using the framework of a linear model for elementary estimates. It easily follows that if *RGB* is negligible, *ABK* is a less efficient alternative to *AK*. Both cases of univariate and multivariate composite estimator are considered. In particular, we consider the question of loss in precision in using *ABK* over *AK* for both univariate and multivariate cases. Some numerical results are presented using the monthly Canadian Labour Force Survey (LFS); the LFS consists of six panels (or rotation groups) based on a multistage stratified cluster design; each panel stays in the sample for six consecutive months so that the month-to-month overlap between samples is 5/6. Note that in the context of the LFS, the matched subsample *m* can be denoted as (-1), *i.e.*, without the birth panel where the number 1 denotes the time-in-sample, *i.e*, 1 month. Similarly, m' can be denoted as (-6), *i.e.*, without the death panel where the number 6 denotes the time-in-sample, *i.e.*, 6 months. Thus, *AK* (*y*) for the LFS is given by

$$AK(y) = F_y + K\left[(F_y^{\prime}-F_y)-(P_{y(-6)}^{\prime}-P_{y(-1)})\right] + A(F_y - P_{y(-1)}), \quad (1.3a)$$

$$= F_y + K(F_y^{\prime}-P_{y(-6)}^{\prime}) + (A-K)(F_y - P_{y(-1)}). \quad (1.3b)$$

The organization of this paper is as follows. Section 2 contains a motivation for the proposed method while Section 3 presents its description. Numerical results are presented in Section 4, and a discussion in Section 5.

2. MOTIVATION

For the sake of motivating the proposed estimator, we first assume that there is no *RGB*, and then give an alternative derivation of (univariate and multivariate) *AK*-estimators using the linear zero function method of Rao (1968) for linear models. The presentation is in the context of the LFS rotation pattern although the basic approach is applicable in general to other rotation patterns.

2.1 Univariate Case

Denote the six panel level *GR* estimates for the previous month $t^{\prime}$ as $f_{y(1)},\dots,f_{y(6)}$ and for t as $g_{y(1)},\dots,g_{y(6)}$ where the number in the subscript signifies time-in-sample. Thus, the five pairs $(f_{y(i)}, g_{y(i+1)})$, $i=1,\dots,5$, represent common panels over $t^{\prime}$ and t, $f_{y(6)}$ is for the death (or rotate-out) panel at $t^{\prime}$ and $g_{y(1)}$ is for the birth (or rotate-in) panel at *t*. Now, consider the following finite population (semi parametric) model.

$$\begin{pmatrix} f_{y(1)} \\ g_{y(2)} \\ \dots \\ f_{y(2)} \\ g_{y(3)} \\ \dots \\ . \\ . \\ . \\ \dots \\ f_{y(6)} \\ g_{y(1)} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ \dots & \dots \\ 1 & 0 \\ 0 & 1 \\ \dots & \dots \\ . & \\ . & \\ . & \\ \dots & \dots \\ 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \theta_{y^{\prime}t^{\prime}} \\ \theta_{yt} \end{pmatrix} + \varepsilon_y , \quad (2.1)$$

where ε_y's have mean zero and covariances given by

$$\text{cov}\ (f_{y(i)}, f_{y(j)}) = \sigma^2_{y^{\prime}t^{\prime}}\ i=j\ ;\ 0 \text{ if } i \neq j \quad (2.2a)$$

$$\text{cov}\ (g_{y(i)}, g_{y(j)}) = \sigma^2_{yt}\ i=j;\ 0 \text{ if } i \neq j \quad (2.2b)$$

$$\text{cov}(f_{y(i)}, g_{y(i+1)}) = \varphi_{y1}\sigma_{y^{\prime}t^{\prime}}\sigma_{yt},\ i=1,\dots,5;$$
$$\varphi_{y2}\sigma_{y^{\prime}t^{\prime}}\sigma_{yt} \text{ for } i=6,7(\text{mod}6). \quad (2.2c)$$

Let $g_y = \Sigma_{i=1}^{6} g_{y(i)}$. Thus $g_y/6$ is the usual *GR*-estimate at *t*. Similarly, let $f_y = \Sigma_{i=1}^{6} f_{y(i)}$. Also let $g_{y(-j)} = g_y - g_{yj}$, and $f_{y(-j)} = f_y - f_{yj}$. It follows that $g_y = \Sigma_{j=1}^{6} g_{y(-j)}/5$, $f_y = \Sigma_{j=1}^{6} f_{y(-j)}/5$.

Now, the BLUE (best linear unbiased estimate) of θ_{yt} can be obtained as an optimal linear combination of the usual estimator $g_y/6$ and all the linearly independent (10 in this case) unbiased estimates of 0, viz., $g_{y(-j)}/5 - g_y/6$, $f_{y(-j)}/5 - f_y/6$ for $j=1,\dots,5$. Thus $\hat{\theta}^{BLUE}_{yt}$ is given by the residual of the regression of $g_y/6$ on the ten zero functions, which is equivalent to the residual of the regression of $g_y/6$ on $g_{y(-\overline{j+1})}/5 - g_{y(-j)}/5$, $f_{y(-\overline{j+1})}/5 - f_{y(-j)}/5, j=1,2,3,4$, and $g_{y(-6)}/5 - g_y/6$, and $f_{y(-6)}/5 - f_y/6$. Notice also that

$$\text{Cov}[g_y/6, (g_{y(-j)}/5 - g_y/6)] = 0\ j=1,\dots,6, \quad (2.3a)$$

$$\text{Cov}[g_y/6, (f_{y(-j)}/5 - f_y/6)] = -\frac{1}{6}(\varphi_{y1}-\varphi_{y2})\sigma_{yt}\sigma_{yt^{\prime}}/30,\ j \neq 6;$$
$$\frac{1}{6}(\varphi_{y1}-\varphi_{y2})\sigma_{yt}\sigma_{y^{\prime}t^{\prime}}/6,\ j=6\ , \quad (2.3b)$$

$$\text{Cov}[(g_{y(-6)}/5 - g_y/6), (g_{y(-\overline{j+1})}/5 - g_{y(-j)}/5)] = 0, \quad (2.3c)$$

$$\text{Cov}[(g_{y(-6)}/5 - g_y/6), (f_{y(-\overline{j+1})}/5 - f_{y(-j)}/5)] = 0, \text{ for } j=1,\dots,4. \quad (2.3d)$$

Next, using the Gram-Schmidt orthogonalization of the regressors, it follows from (2.3) above that it is enough to retain $g_{y(-6)}/5 - g_y/6$ and $f_{y(-6)}/5 - f_y/6$ as regressors. Thus, $\hat{\theta}^{BLUE}_{yt}$ is the best linear combination of $g_y/6$ (or F_y), $(f_{y(-6)}/5 - f_y/6)$ (or $P^{\prime}_{y(-6)} - F^{\prime}_y$), and $(g_{y(-6)}/5 - g_y/6)$ (or $P_{y(-6)} - F_y$), which coincides with the AK-estimator given in (1.3b).

We will use the term "primary predictor" for $(f_{y(-6)}/5 - f_y/6)$ as it is correlated with the initial estimator $g_y/6$, and the term "secondary predictor" for $(g_{y(-6)}/5 - g_y/6)$ as it is uncorrelated with $g_y/6$. The notation AK^* will denote the AK-estimator with only the primary predictor.

2.2 Multivariate Case

Without loss of generality, suppose there are two study variables y and z, and the parameter vector of interest is $\theta_t = (\theta_{yt}, \theta_{zt})'$. Now denoting by $\boldsymbol{f}_j$ the vector of panel-level estimates $(f_{yj}, f_{zj})'$ and $\boldsymbol{g}_j$ in a similar manner, it can be easily shown as in the univariate case that the BLUE of $\boldsymbol{\theta}_t$ is given by the optimal linear combination of the vectors $\boldsymbol{g}/6, \boldsymbol{f}_{(-6)}/5 - \boldsymbol{f}/6$, and $\boldsymbol{g}_{(-6)}/5 - \boldsymbol{g}/6$. Again $\boldsymbol{f}_{(-6)}/5 - \boldsymbol{f}/6$ serves as the primary predictor vector and $\boldsymbol{g}_{(-6)}/5 - \boldsymbol{g}/6$ as the secondary. Thus, $\hat{\theta}_t^{BLUE}$ is a natural multivariate analogue of the AK-estimator, denoted by m-AK, and is given by

$$\begin{pmatrix} m\text{-}AK(y) \\ m\text{-}AK(z) \end{pmatrix} = \begin{pmatrix} F_y \\ F_z \end{pmatrix} + \boldsymbol{K} \begin{pmatrix} F_y' - P_{m(y)}' \\ F_z' - P_{m(z)}' \end{pmatrix} + (\boldsymbol{A} - \boldsymbol{K}) \begin{pmatrix} F_y - P_{m(y)} \\ F_z - P_{m(z)} \end{pmatrix}, \quad (2.4)$$

where the coefficients $\boldsymbol{K}, \boldsymbol{A}$ now denote the optimal 2x2 matrices for the bivariate case. With only primary predictors, the estimator will be denoted as m-AK^*.

3. PROPOSED METHOD IN THE PRESENCE OF *RGB*

3.1 Univariate Case

In the presence of *RGB*, the 10 predictor zero functions used earlier in Section 2.1 are no longer zero functions. This is so because in the model (2.1), the means of $f_{y(j)}$ and $g_{y(j)}$ respectively are now $\theta_{yt'} + b_{yj}$ and $\theta_{yt} + b_{yt}, j = 1, \ldots, 6$ where b_{yj} denotes the time-stationary bias for the j-th month-in-sample panel for the characteristic y. Note that $\Sigma_{j=1}^{6} b_{yj} = 0$. Now observe that $g_{y(-j)}/5$ and $f_{y(-j)}/5$ estimate respectively $\theta_{yt} + b_{yj}/5$ and $\theta_{yt'} + b_{yj}/5$. We therefore have six unbiased estimates of the difference $\theta_{yt} - \theta_{y't'}$ given by $g_{y(-j)}/5 - f_{y(-j)}/5$ for $j = 1, \ldots, 6$. These in turn give six estimates of 0, denoted by $h_{y(-j)}$ and defined as, for $j = 1, \ldots, 6$,

$$h_{y(-j)} := (g_{y(-j)} - f_{y(-j)})/5 - (g_y - f_y)/6 \quad (3.1)$$

Notice that only five of the above six $h_{y(-j)}$'s are linearly independent in view of the constraint $\Sigma_{j=1}^{6} h_{y(-j)} = 0$.

The next step is to find the best linear combination of $g_y/6$ and the five zero functions $h_{y(-j)}$. It follows from the Appendix that the BLUE of θ_{yt} is given by the best linear combination of $g_y/6$ and the three predictor zero functions $h_{y(6)}, h_{y(-1)} - (h_{y(-2)} + h_{y(-3)} + h_{y(-4)})/3$, and $h_{y(-5)} - (h_{y(-2)} + h_{y(-3)} + h_{y(-4)})/3$. The predictor $h_{y(-6)}$ is primary while the other two are secondary. The predictors can be further simplified to $(P_{y(-6)} - P_{y(-6)}') - (F_y - F_y')$, and $(P_{y(1)} - P_{y(1)}') - (P_{y(2,3,4)} - P_{y(2,3,4)}')$, and $(P_{y(5)} - P_{y(5)}') - (P_{y(2,3,4)} - P_{y(2,3,4)}')$ where $P_{y(2,3,4)}$, *e.g.*, denotes the partial sample estimate based on panels with months-in-sample 2, 3, and 4. Thus, the univariate ABK-estimator (with three coefficients K, A, and B corresponding to the three predictors) is given by

$$\begin{aligned} u\text{-}ABK(y) = F_y + K[(F_y - F_y') - (P_{y(-6)} - P_{y(-6)}')] + \\ B[(P_{y(2,3,4)} - P_{y(2,3,4)}') - (P_{y(1)} - P_{y(1)}')] + \\ A[(P_{y(2,3,4)} - P_{y(2,3,4)}') - (P_{y(5)} - P_{y(5)}')]. \quad (3.2) \end{aligned}$$

Note that the predictors for ABK are different from those for AK. The primary predictor involves the difference of two estimates of $\theta_{yt} - \theta_{y't'}$, one based on the full sample and the other based on panels not including the death panel (*i.e.*, with subscript 6). The two secondary predictors also involve differences of two estimates of $\theta_{yt} - \theta_{y't'}$. One secondary predictor is based on the difference for birth panel (*i.e.*, the panel succeeding the death panel) and that for the remaining panels not including the death panel or the panels succeeding or preceding it. The other secondary predictor is based on the difference for the panel preceding the death panel and that for the remaining panels as in the case of the first secondary predictor. As before, the u-ABK-estimator with only the primary predictor will be denoted by u-ABK^*.

3.2 Multivariate Case

The multivariate case in the presence of *RGB* is analogous to the m-AK estimator. The coefficients K, A, B become matrices. In the bivariate case, for example, m-ABK takes the form

$$\begin{pmatrix} m\text{-}ABK(y) \\ m\text{-}ABK(z) \end{pmatrix} = \begin{pmatrix} F_y \\ F_z \end{pmatrix} + K\begin{pmatrix} (F_y - F_y') - (P_{y(-6)} - P_{y(-6)}') \\ (F_z - F_z') - (P_{z(-6)} - P_{z(-6)}') \end{pmatrix} + B\begin{pmatrix} (P_{y(2,3,4)}) - (P_{y(2,3,4)}') - (P_{y(1)} - P_{y(1)}') \\ (P_{z(2,3,4)} - P_{z(2,3,4)}') - (P_{z(1)} - P_{z(1)}') \end{pmatrix} + A\begin{pmatrix} (P_{y(2,3,4)} - P_{y(2,3,4)}') - (P_{y(5)} - P_{y(5)}') \\ (P_{z(2,3,4)} - P_{z(2,3,4)}') - (P_{z(5)} - P_{z(5)}') \end{pmatrix} \quad (3.3)$$

If only the primary predictors (*i.e.*., those with the coefficient matrix K) are used, then the estimator will be denoted as *m-ABK** as before.

3. 3 Comparison of *ABK* with *AK*

If *RGB* is absent, *i.e.*, $b_{y(j)}=0$ for all *j*, then the underlying semiparametric model (2.1) is the same for both *AK* and *ABK*. Therefore, since both *AK* and *ABK* are unbiased, and *AK* is BLUE, clearly *ABK* will be less efficient than *AK*.. However, if $RGB \neq 0$, the model (2.1) no longer holds, and the *AK*-estimator will be biased, whereas *ABK* is not only unbiased but also optimal in the class of all unbiased estimates. It follows that in the presence of *RGB*, there is a trade-off between bias and variance as one moves from *AK* to *ABK*.

4. NUMERICAL EXAMPLES

The main purpose of the numerical study was to get some idea of the magnitude of loss in efficiency in using *ABK* over *AK*. For computational simplicity, we compared only *ABK** and *AK** although they are not fully optimal as they use only primary predictors. Nevertheless, this may be adequate for understanding loss in efficiency. Also for computational simplicity, we considered composite estimates using information from only two consecutive time points. Also, efficiencies for several pairs of consecutive time points were averaged. Note that this averaging of efficiencies is justifiable because they are all based on the same number of time points, *i.e.*, two.

The Ontario LFS data for 1991 was used for the study. The variables studied were employment (E) and unemployment (U). For the multivariate case, eight variables were used which are defined by cross-identifying (E, U) by gender by age (15 - 24, 25+). The optimal coefficients for *AK** and *ABK** were computed adaptively for each time-point via a jackknife estimate of the underlying covariance matrix; the degrees of freedom for this estimate of the covariance matrix can be roughly measured in terms of the number of pseudo-replicates minus the number of strata which is about 1500 minus 300. These estimated optimal coefficients were, however, treated as fixed in estimating the variance of the composite estimate. This may be reasonable for the purpose of evaluating the efficiency of *ABK* relative to *AK*. Note that the usual method of computing optimal coefficients for *AK* is quite tedious. It entails first finding a smoothed estimate of the underlying covariance matrix under some simplifying assumptions, and then a grid search is used to find the optimal pair (*A*, *K*). For the multivariate case, this can be very complex.

Table 1 shows the average gain in efficiency of *AK** over *GR* and the reduction in efficiency gain of *ABK** relative to the gain due to *AK** for the period of 11 months (Feb-Dec '91). For example, the 5.15% gain in E when *AK** is used becomes a 3.08% gain; when *ABK** is used, a reduction of 40%. Notice that the gains in efficiency of *AK** are not high because we are not using all the *AK*-predictors as well as not all the information from the past except for the previous time point. It is observed from Table 1 that the reduction in efficiency gain due to the use of *ABK** may be quite substantial. It ranges from 9% to 40% approximately.

Table 1

Average Efficiency Gains for *AK and *ABK***

	Efficiency Gain of *AK**		Reduction in Efficiency Gain of *ABK**	
	E	U	E	U
Univariate	5.15 %	3.53 %	40.19 %	31.16 %
Multivariate	6.51 %	4.8 %	30.72%	8.96 %

Note: The superscript "*" indicates that only primary predictors were used in composite estimation.

5. DISCUSSION

Using the theory of optimal linear regression, a composite estimator (*ABK*) in the presence of rotation group bias was proposed. The multivariate case was also included in which correlation between selected study variables for a given time point was also exploited. The *RGB* was modelled under the usual conditions. A numerical example from the LFS was also presented. An interesting observation based on the numerical example was that there could be substantial reduction in efficiency gains when one moves from *AK* (composite estimator in the absence of *RGB*) to *ABK*. This suggests that it may be preferable to take measures to control sources of *RGB* so that *RGB* would be deemed negligible, such as improved interviewer training, and panel-level nonresponse adjustment.

Despite the reduced level of efficiency gains, *ABK* does provide a solution to adjust *AK* for *RGB* under fairly general conditions. The rotation pattern used for defining *ABK* was based on the LFS. However, the underlying idea is quite general, and can be used for alternative patterns; for instance, the rotation pattern for the U.S. Current Population Survey (CPS) is 4-8-4, *i.e.*, a panel stays in the sample for four months, temporarily rotates out for eight months, and then rotates in for four months. Thus, every month consist of eight panels with month-in-sample 1, 2, ... , 8, out of which {1,5} correspond to "birth" panels (in the sense of first time, and fifth time after eight months of break), {4, 8} correspond to "death" panels, {3, 7} correspond to panels preceding the death panels (as opposed to birth panels which succeed the death panels), and {2, 6} correspond to the remaining panels, *i.e.*, neither succeeding nor preceding the death panels. It follows from (3.2), that the *u-ABK* estimator for the 4-8-4 rotation pattern under the usual linear model assumptions for the elementary estimates is given by

$$u\text{-}ABK(y) = F_y + K[(F_y - F'_y) - (P_{y(-4,-8)} - P'_{y(-4,-8)})]$$
$$+ B[(P_{y(2,6)} - P'_{y(2,6)}) - (P_{y(1,5)} - P'_{y(1,5)})]$$
$$+ A[(P_{y(2,6)} - P'_{y(2,6)}) - (P_{y(3,7)} - P'_{y(3,7)})] \quad (5.1)$$

The multivariate version can also be defined in a similar manner.

APPENDIX

We need the following lemma.

Lemma

(a) $\mathrm{Cov}(g_y/6, h_{y(-j)} - h_{y(-j')}) = 0 \ , j, j' = 1, \ldots, 5$

(b) $\mathrm{Cov}(h_{y(-6)}, h_{y(-j)} - h_{y(-j')}) = 0, \ j, j' = 2,3,4;$
$\neq 0, j = 1,5, j' = 2,3,4.$

(c) $\mathrm{Cov}(h_{y(-j)}, h_{y(-2)} - h_{y(-3)}) = 0$
$= \mathrm{Cov}(h_{y(-j)}, h_{y(-3)} - h_{y(-4)}), \ j = 1,5$

Proof:

(a) follows easily by noting that

$$\mathrm{Cov}(g_y/6, h_{y(-j)}) = \frac{1}{30}(\varphi_{y1} - \varphi_{y2})\sigma_{yt}\sigma_{yt'}/6, j = 1 \text{ to } 5;$$
$$-\frac{1}{6}(\varphi_{y1} - \varphi_{y2})\sigma_{yt}\sigma_{yt'}/6, \ j = 6.$$

(b) follows from the observation that

$$\mathrm{Cov}(h_{y(-6)}, h_{y(-j)}) =$$
$$-\frac{1}{30}\left[\frac{\sigma^2_{yt} + \sigma^2_{y't'}}{5} - \frac{1}{15}(4\varphi_{y1} + 2\varphi_{y2})\sigma_{yt}\sigma_{y't'}\right], j = 2,3,4;$$
$$-\frac{1}{30}\left[\frac{\sigma^2_{yt} + \sigma^2_{y't'}}{5} + \frac{1}{15}(14\varphi_{y1} - 2\varphi_{y2})\sigma_{yt}\sigma_{y't'}\right], j = 5;$$
$$-\frac{1}{30}\left[\frac{\sigma^2_{yt} + \sigma^2_{y't'}}{5} - \frac{1}{15}(\varphi_{y1} - 13\varphi_{y2})\sigma_{yt}\sigma_{y't'}\right], j = 1.$$

(c) follows from the fact that

$$\mathrm{Cov}(h_{y(-j)}, h_{y(-j')})$$
$$= -\frac{1}{150}[(\sigma^2_{yt} + \sigma^2_{y't'}) + 4(2\varphi_{y1} - \varphi_{y2})\sigma_{yt}\sigma_{y't'}],$$
$$\text{for } j = 1,5; j' = 2,3,4.$$

Now, *u-ABK* (y) is given by the optimal linear combination of $g_y/6$, and any five linearly independent functions of $h_{y(-1)}, \ldots, h_{y(-6)}$ such as $h_{y(-6)}, h_{y(-1)}, h_{y(-5)}, h_{y(-2)} - h_{y(-3)}, h_{y(-3)} - h_{y(-4)}$. Further, from the lemma since $g_y/6, h_{y(-6)}, h_{y(-1)}$, and $h_{y(-5)}$ are orthogonal to $h_{y(-2)} - h_{y(-3)}$ and $h_{y(-3)} - h_{y(-4)}$,

u-$ABK(y)$ is simply the optimal linear combination of $g_y/6$, $h_{y(-6)}$, $h_{y(-1)}$, and $h_{y(-5)}$; or equivalently $g_y/6$, $h_{y(-6)}$, $h_{y(-1)} + (h_{y(-6)} + h_{y(-1)} + h_{y(-5)})/3$, and $h_{y(-5)} + (h_{y(-6)} + h_{y(-1)} + h_{y(-5)})/3$, which, in turn, reduces to $g_y/6$, $h_{y(-6)}$, $h_{y(-1)} - (h_{y(-2)} + h_{y(-3)} + h_{y(-4)})/3$, $h_{y(-5)} - (h_{y(-2)} + h_{y(-3)} + h_{y(-4)})/3$, using the fact that $\sum_{j=1}^{6} h_{y(-j)} = 0$. With this set of three predictor zero functions, it is easy to see from the above lemma that $g_y/6$ is orthogonal to the two predictors $h_{y(-1)} - (h_{y(-2)} + h_{y(-3)} + h_{y(-4)})/3$, $h_{y(-5)} - (h_{y(-2)} + h_{y(-3)} + h_{y(-4)})/3$, but $h_{y(-6)}$ is not orthogonal to these two predictors. Thus, the predictor $h_{y(-6)}$ is primary while the other two predictors are secondary for the purpose of improving the initial estimator $g_y/6$.

The formula for *m-ABK(y)* can be similarly established.

ACKNOWLEDGEMENT

The authors are grateful to Jon Rao and Harold Mantel for helpful discussions. The first author's research was supported in part by a grant from Natural Sciences and Engineering Council of Canada held at Carleton University, Ottawa under an adjunct research professorship.

REFERENCES

Bailar, B.A. (1975). "The effects of rotation group bias on estimates from panel surveys", *Journal of the American Statistical Association*, 70, 23-29.

Fuller, W.A. (1990). "Analysis of repeated surveys", *Survey Methodology*, 16, 167-180.

Gurney, M., and Daly, J.F. (1965). "A multivariate approach to estimation in periodic sample surveys", *Proceedings of the American Statistical Association, Social Statistics Section*, 242-257.

Hansen, M.H., Hurwitz, W.N., and Madow, W.G. (1953). *Sample Survey Methods and Theory*, Vol. 2, New York: Wiley.

Huang, E.T., and Ernst, L.R. (1981). "Comparison of an alternative estimator to the current composite estimator of CPS", *Proceedings of the American Statistical Association, Section on Survey Research Methods*, 303-308.

Kumar, S., and Lee, H. (1983). "Evaluation of composite estimation for the Canadian Labour Force Survey", *Survey Methodology*, 9, 1-24.

Rao, C.R. (1968). "A note on a previous lemma in the theory of least squares, and some further results". *Sankhyā, A30*, 259-266.

Smith, T.M.F. (1978). "Principles and problems in the analysis of repeated surveys", in: N. Krishan Namboodiri, ed., *Survey Sampling Measurement*, New York: Academic Press, 201-206.

Valliant, R. (1993). "Poststratification and conditional variance estimation", *Journal of the American Statistical Association, 88*, 89-96.

Wolter, K.M. (1979). "Composite Estimation in finite populations", *Journal of the American Statistical Association, 74*, 604-613.