COMPOSITE ESTIMATION BY MODIFIED REGRESSION FOR REPEATED SURVEYS

A.C. Singh and P. Merkouris
A.C. Singh, Survey and Analysis Methods Development, Statistics Canada, 16A Coats Bldg, Ottawa K1A 0T6

1. INTRODUCTION

It is a common practice for statistical agencies to conduct large scale surveys which are repeated over time and consist of overlapping panels. For example, the monthly Canadian Labour Force Survey (LFS) consists of six panels (or rotation groups) based on a multistage stratified cluster design for the non-institutional population of age 15 or over. In LFS, each panel stays in the sample for six consecutive months, so that the month-to month overlap between samples is 5/6. The main characteristics or study variables of interest are monthly levels, month-to-month changes and quarterly averages for employment (E) and unemployment (U). In this paper, we will mainly be concerned with estimation of levels for broad (in contrast to small) areas. From these estimates, month-to-month changes and quarterly averages can of course be obtained.

Since several study variables are measured on the same unit, the estimate of a variable is correlated with estimates of other variables. Furthermore, due to overlapping samples, the estimate of a variable is also correlated with estimates of the same variable , as well as other variables, over time. Clearly, the precision of the usual estimate of a variable can be improved, in general, by using estimates of other correlated variables. It is interesting to note that the provision of partial overlap is not only attractive from the view-point of saving field costs and reducing respondent burden, but also essential for obtaining more precise estimates using past information. If there is complete overlap or no overlap, no improvement can be made. Throughout this article, we follow the d-based (d for design) approach in which finite population parameters are assumed nonrandom. The alternative model-based approach involving modelling of random finite population parameters (as in the case of small area estimation) will not be considered here. Incidentally, in the model-based approach partial overlap is not necessarily required.

The term "composite estimate" will be used in general to denote a linear combination of current and correlated past estimates adjusted for change. The prefix "univariate" will be used when current and past information on the same variable is used. The prefix "multivariate" will be used when current and past information about several variables is used. More specifically, the set-up can be as follows. Consider two occasions $t' < t$, for simplicity, and let $\hat{\theta}_{y't'}$ and $\hat{\theta}_{yt}$ denote respectively the usual (full-sample) estimates of the finite population parameters (e.g. totals) $\theta_{y't'}$ and θ_{yt} for variables y' at t' and y at t; $\hat{\theta}_{yt}$ is typically the generalized regression (GR) estimator which incorporates information about auxiliary variables (such as demographic) with known population totals. Also, let $\hat{\theta}_{ytm}$ and $\hat{\theta}_{y't'm'}$ denote the corresponding estimates for the matched subsample; m signifies (backward) matching with respect to t', and m' signifies (forward) matching with respect to t. Now, with two occasions, there are three estimates for θ_{yt}. These are $\hat{\theta}_{yt}$, $\hat{\theta}_{ytm}$ and $\hat{\theta}_{y't'}$ adjusted for change, i.e., $\hat{\theta}_{y't'} + (\hat{\theta}_{ytm} - \hat{\theta}_{y't'm'})$. The problem of (univariate) composite estimation is to find a suitable linear combination of these three estimates. If z denotes another correlated variable at time t (e.g., z could be E and y could be U in the case of LFS), then the problem of (multivariate) composite estimation is to find a suitable linear combination of five estimates: $\hat{\theta}_{yt}$, $\hat{\theta}_{ytm}$, $\hat{\theta}_{y't'} + (\hat{\theta}_{ytm} - \hat{\theta}_{y't'm'})$, $\hat{\theta}_{zt} + (\hat{\theta}_{ytm} - \hat{\theta}_{ztm})$ and $\hat{\theta}_{z't'} + (\hat{\theta}_{ytm} - \hat{\theta}_{z't'm'})$.

The literature on composite estimation is quite rich and covers a period of over fifty years. For a review, see Binder and Hidiroglou (1988). For finite populations, an estimator termed K-Composite was proposed by Hansen, Hurvitz, and Madow (1953), and its properties were studied by Rao and Graham (1964) for general rotation schemes. Gurney and Daly (1965) proposed a more general estimator termed AK-Composite based on elementary or panel level estimates. In Section 2, we review K- and AK-Composite estimators as well as their limitations which explain why they are not commonly used. To get around some of these limitations, recently Fuller (1990) proposed an important idea of composite weighting (referred to in this paper as AKC, C for calibrated) which relates composite estimation to weight calibration in sampling (WCS). The AKC-Composite was studied by Lent, Miller, and Cantwell

(1994). It will be seen in Section 2 that this method is still not satisfactory.

In this paper, we propose a method termed MR-Composite which based on the modified regression methodology of Singh (1994). It is multivariate in nature and overcomes several limitations of existing methods. Like AKC, it also uses WCS for computation, but in a somewhat different manner, and is quite easy to implement. The MR methodology is based on the theory of generalized estimating functions which uses the concept of "working" covariance and provides a simple alternative to the model-assisted approach used in GR. MR extends GR to the general case of correlated auxiliary information from past data in a very natural way. Section 3 contains a review of MR, and Section 4 contains the proposed method of MR-Composite. Empirical results on its performance using LFS data are given in Section 5. The final Section 6 contains concluding remarks.

2. REVIEW OF EXISTING METHODS

All existing methods give rise to regression-type estimators, which are asymptotically optimal under suitable conditions. Only two time points t' and t will be considered in view of the recursive nature of all estimators.

2.1 AK-Composite

This is due to Gurney and Daly (1965). For given constants A, K ($0 < A < K < 1$, in general), it is given by

$$\hat{\theta}_{yt}^{AK} = \hat{\theta}_{yt}^{GR} - (K-A)(\hat{\theta}_{yt}^{GR} - \hat{\theta}_{ytm}^{GR}) + K(\hat{\theta}_{y't'}^{AK} - \hat{\theta}_{y't'm'}^{GR}) \qquad (2.1)$$

$$= (1-K)\hat{\theta}_{yt}^{GR} + A(\hat{\theta}_{yt}^{GR} - \hat{\theta}_{ytm}^{GR}) + K(\hat{\theta}_{y't'}^{AK} + \hat{\theta}_{ytm}^{GR} - \hat{\theta}_{y't'm'}^{GR})$$

The K-Composite of Hansen, Hurvitz and Madow (1953) is a special case with A = 0. The optimal values of A and K, in the sense of minimum variance of the estimator, are in practice obtained by averaging estimated values using data over several time points under the assumption of stationarity.

Note that the term $A(\hat{\theta}_{yt}^{GR} - \hat{\theta}_{ytm}^{GR})$ in the expression (2.1) can also be viewed as reduction of the rotation group bias introduced by the term $K(\hat{\theta}_{yt}^{GR} - \hat{\theta}_{ytm}^{GR})$ in the K-Composite. Rotation group bias implies that individual panel means depend on time-in-sample, and therefore vary from rotation group to rotation group. It is known that the birth panel usually differs most from the others.

2.1.1 Limitations of AK-Composite.

These are listed below.

(a) The optimal choice of (A,K) depends on the study variable. This, in turn, leads to internal inconsistency with respect to the selected variables for composite estimation. This means that component level estimates do not add up to the aggregate level estimate, e.g., for LFS, estimates for E and U may not add up to that for the labour force. This problem could be resolved by using the convention that the estimate for an aggregate would be constructed by adding its components. However, the resulting estimate for the aggregate would not be AK-Composite.

(b) Often the study variables are linearly dependent, e.g., an individual in LFS must have one of the three statuses: E, U, and not in the labour force (N). Any external constraint on the total LFS population count must therefore be satisfied by the three estimates. One could resolve this problem by arbitrarily leaving one variable out from the AK-Composite estimation and compute the corresponding estimate as a residual. However, this solution is not symmetric with respect to the choice of the left-over variable. Moreover, the resulting estimator will not be AK-Composite.

(c) The estimate of (A,K) for each study variable is obtained from covariance matrices from past data under the assumption of stationarity. This assumption may not be reasonable. Moreover, this may be computationally tedious if there are many study variables and the design is complex .

(d) AK-Composite is univariate in nature and thus may yield only marginal gains in efficiency for some variables. In principle, it can be made multivariate for further gains in efficiency. However, this will make computation of coefficients for the linear combination even harder and their estimates may not be stable due to insufficient degrees of freedom.

(e) There is another type of internal inconsistency. For variables not originally planned for composite estimation, finding corresponding (A,K) will not only be cumbersome, it may not even be feasible to ensure that the resulting estimates remain internally consistent with the AK-Composite estimates for the planned variables. This is a rather serious limitation.

In view of these limitations, AK-Composite is not commonly used. An important exception is the U.S. Current Population Survey which uses a compromise choice of (A,K) as (.2, .4) for the main three labour force variables, the choice being approximately optimal for estimates of unemployment level.

2.2 AKC-Composite

The AK-Calibration composite estimation overcomes the serious limitation (e) of AK-Composite and is due to Fuller (1990). It was studied by Lent, Miller, and Cantwell (1994) in the context of CPS. It is based on the idea of estimation via WCS, and thus produces composite weights to be used for estimation of any other variable. It consists of two steps. First, for a few linearly independent key variables (such as E and U for LFS at suitable levels of aggregation), the AK-Composite estimators are obtained. Second, the composite weights are obtained by using these estimates as additional auxiliary population totals in a calibration method such as regression or raking. For LFS, the usual auxiliary totals for GR computed via WCS correspond to demographic counts for domains defined by age, sex and region. Note that for the key variables used as auxiliary totals, the AKC-Composite will coincide with the AK-Composite by construction. For other variables, the AKC-Composite is not optimal but is expected to yield efficiency gains over the usual estimates because of correlation of these variables with the key variables. Also note that the limitation (e) of AK-Composite no longer exists because all the estimates are built from common building blocks. However, AKC-Composite still inherits all the other limitations (a) to (d) of AK.

3. HEURISTIC MOTIVATION OF MR

The AK-Composite is motivated from d-based linear models for elementary estimates. It is known that Godambe and Thompson's (1989) method of estimating function (MEF), and the generalized estimating function (GEF) when the true covariance matrix is replaced by a working covariance, provide general alternatives to linear modelling. We show that GEF gives insight into overcoming limitations of AK-Composite. For this purpose, a d-based GEF is needed for finite population parameters. This was introduced by Singh (1994) who used this to propose a modified regression (MR) method for linear finite population parameters. To motivate the proposed method we will review d-GEF via MR in the context of repeated surveys.

The MR method uses pieces of information from the data in the form of d-based zero functions (i.e., functions which are zero in expectation). These functions depend, in general, on both data and parameters. However, they may be parameter-free provided they are correlated with other zero functions. Now for convenience, consider the case of MR for a single time point and only one parameter θ_{yt}. Similar to the concept of elementary estimates in d-based linear models, consider zero functions based on Horvitz-Thompson (HT) estimates. For example, suppose for time t there are p-auxiliary variables $\boldsymbol{x}_t$ with known population totals and the study variable is y. The zero function involving θ_{yt} is $\hat{\theta}_{yt}^{HT}-\theta_{yt}$ or $\sum_{k\in s(t)}h_{kt}y_{kt}-\theta_{yt}$, where h_{kt} is the inverse of the selection probability for the unit k, and s(t) denotes the sample at t with size n_t. This function will be denoted as $\sum_{k\in U(t)}\phi_{kt}h_{kt}$ where $\phi_{kt}=(1_{k\in s(t)}-h_{kt}^{-1})y_{kt}$, $1_{k\in s(t)}$ is the indicator function and $U(t)$ is the finite population at t. The other zero functions do not involve the parameter and correspond to auxiliary variables. The function for the ith x-variable, $\hat{\theta}_{xti}^{HT}-\theta_{xti}$, will be denoted as $\sum_{k\in U(t)}\psi_{kti}h_{kt}$ where $\psi_{kti}=(1_{k\in s(t)}-h_{kt}^{-1})x_{kti}$.

The MR for θ_{yt} based on data for single time t is given as a solution of the estimating equation $\boldsymbol{G}_t'\boldsymbol{\Gamma}_t^{-1}\boldsymbol{g}_t=\boldsymbol{0}$ where $\boldsymbol{g}_t$ is the $(p+1)$-vector $(\sum\phi_{kt}h_{kt},\sum\psi_{kt1}h_{kt},...,\sum\psi_{ktp}h_{kt})'$, $\boldsymbol{\Gamma}_t$ is a $(p+1)\times(p+1)$ working d-based covariance matrix of $\boldsymbol{g}_t$, and $\boldsymbol{G}_t$ is the $(p+1)\times 1$ matrix $(-1,0,...,0)'$. If $\boldsymbol{\Gamma}_t$ is the true covariance matrix, then the estimating function, and hence the estimator, becomes optimal in a certain sense. However, with a working choice of $\boldsymbol{\Gamma}_t$, the resulting estimator becomes suboptimal, but robust in that it continues to be asymptotically design consistent. One possible choice of $\boldsymbol{\Gamma}_t$ is

$$\boldsymbol{\Gamma}_t=\begin{bmatrix}\tilde{\boldsymbol{\phi}}_t'\boldsymbol{\Lambda}_t\tilde{\boldsymbol{\phi}}_t & \tilde{\boldsymbol{\phi}}_t'\boldsymbol{\Lambda}_t\tilde{\boldsymbol{\Psi}}_t\\ - & \tilde{\boldsymbol{\Psi}}_t'\boldsymbol{\Lambda}_t\tilde{\boldsymbol{\Psi}}_t\end{bmatrix} \tag{3.1}$$

where the superscript " ~ " denotes only the part which involves the random variable $1_{k\in s(t)}$ and is nonzero for the observed sample. Thus, $\tilde{\boldsymbol{\phi}}_t$ is the n_t-vector of $\tilde{\phi}_{kt}$-values, $\tilde{\boldsymbol{\Psi}}_t$ is a $n_t\times p$ matrix whose ith column is the n_t-vector of $\tilde{\psi}_{kti}$-values, and $\boldsymbol{\Lambda}_t$ is the $n_t\times n_t$ matrix diag $(h_{kt},\ 1\leq k\leq n_t)$. With this choice of $\boldsymbol{\Gamma}_t$, $\hat{\theta}_{yt}^{MR}$ in fact coincides with the usual $\hat{\theta}_{yt}^{GR}$, and is given by $\sum_{k\in U(t)}y_{kt}h_{kt}-\hat{\boldsymbol{\beta}}_t'(\sum_{k\in U(t)}\psi_{kt}h_{kt})$, where $\hat{\boldsymbol{\beta}}_t=\boldsymbol{\phi}_t'\boldsymbol{\Lambda}_t\tilde{\boldsymbol{\Psi}}_t'(\tilde{\boldsymbol{\Psi}}_t'\boldsymbol{\Lambda}_t\tilde{\boldsymbol{\Psi}})^{-1}$.

The MR approach gives a general alternative formulation to the traditional model-assisted approach which gives rise to GR. Under MR, since the form of the estimating function is simple, it is easy to incorporate more information via zero functions; this feature will be useful for multivariate composite estimation. The provision of working covariance matrix allows flexibility in dealing with complex ψ-zero functions, e.g., in the case of auxiliary variables with known but random population totals. Furthermore, in MR, since the zero functions are constructed directly, they can be manipulated or transformed to suit the situation; this feature will be especially useful in dealing with correlated past information. To do the corresponding task with GR would require a suitable formulation of the superpopulation regression model which may be tedious. Now, to define MR for two time points t' and t, consider the additional zero functions $\hat{\theta}^{GR}_{ytm} - \hat{\theta}^{GR}_{yt}$, $\hat{\theta}^{GR}_{y't'm'} - \hat{\theta}^{GR}_{y't'}$ ($\hat{\theta}^{GR}_{y't'}$ could be replaced by previous occasion's best estimator obtained recursively). More zero functions corresponding to other study variables, z, could also be defined in a similar manner. All that is required now is to define a suitable working covariance matrix. In the next section, we show how to do this following the above motivation.

4. MR-COMPOSITE: PROPOSED METHOD

4.1 Description

The proposed method departs from the traditional AK-Composite in several ways. We describe the method by explaining these departures. For the recursive set-up, assume that $\hat{\theta}^{MR}_{y't'}$ is given.

4.1.1 Transformation of Zero Functions. It follows from the motivation for MR that GR can be easily extended to include additional zero functions (these are of two types: matched subsample minus full-sample estimates for the current and previous time points), if they can be transformed into a ψ-zero function, i.e., as $\hat{\theta}^{HT}_{xt} - \theta_{xt}$ for a suitable choice of x. The zero function for the current time is relatively easy to express as a ψ-zero function. We first transform it to $\hat{\theta}^{HT}_{ytm} - \hat{\theta}^{HT}_{yt}$ and then define ψ-function as $[(\alpha)1_{k\in s(mt)} - 1_{k\in s(t)}]y_{kt}$, where $s(mt)$ is the matched subsample with respect to t', and α is the ratio adjustment factor required to get population total estimate from the matched subsample. The factor α is simply $\sum_{k\in s(t)} h_{kt} / \sum_{k\in s(mt)} h_{kt}$; for LFS it is approximately 6/5. For this zero function, the corresponding known population total is naturally taken as zero. Now for the zero function corresponding to the previous time point, since $s(mt) = s(m't')$, information about y' from t' for the matched subsample can be augmented to the current data by micro-matching. For this purpose, some imputation (such as carrying the current y-value backward) may be needed for missing data due to nonresponse at t' and movers from t' to t. We then express the zero function $\hat{\theta}^{GR}_{y't'm'} - \hat{\theta}^{MR}_{y't'}$ as $\hat{\theta}^{GR}_{y'tm} - \tilde{\theta}^{MR}_{y't}$ where $\tilde{\theta}^{MR}_{y't}$ is $\hat{\theta}^{MR}_{y't'}$ adjusted for change, i.e., $\hat{\theta}^{MR}_{y't'} + (\hat{\theta}^{GR}_{y'tm} - \hat{\theta}^{GR}_{y't'm'})$; this is an unbiased estimate of $\theta_{y't}$, the current population total for the previous occasion's variable. Here we assume, in effect, that information about y' was also collected at t from respondents in the matched portion. Next we regard y' at t as a new auxiliary variable with known (but random) population total $\tilde{\theta}^{MR}_{y't}$, which implies that the transformed zero function $\hat{\theta}^{HT}_{y'tm} - \tilde{\theta}^{MR}_{y't}$ can be expressed as a ψ-zero function. This suggests constructing MR-Composite as a GR-type estimator.

4.1.2 Suboptimal Coefficients. In view of the above modification to the additional zero functions, the GR-type working covariance matrix can be used for defining MR-Composite. In particular, the coefficients in the linear combination for MR-Composite will vary with t, and can be obtained adaptively from the current and previous occasions' samples. These coefficients will, of course, be only suboptimal unlike the optimal ones used for AK-Composite.

4.1.3 Multivariate in nature. The above GR-formulation of MR-Composite clearly suggests that it can be made multivariate quite easily. Inclusion of more zero functions corresponding to other study variables will not cause any new problems in finding coefficients of the linear combination. All the coefficients can be computed simultaneously in the usual manner.

4.1.4 Subset of Auxiliary Variables. Traditionally, the additional auxiliary variables or zero functions come in pairs; each variable gives rise to two zero functions, one for the current time and other for the previous time point. However, from the efficiency perspective, it is better to have a control on the choice of predictors; too few may not yield enough gain, while too many may cause a loss in precision due to possible instability in estimating regression coefficients. For MR-Composite, it is easy to incorporate this control on predictors.

4.1.5 Computation by WCS. It is clear from the above GR-formulation of MR-Composite that the usual regression-WCS can be used for its computation. This is not only convenient in practice, but also useful for producing estimates for variables not used as additional auxiliary ones. This of departure of MR-Composite from AK-Composite is somewhat similar to that of AKC-Composite. However, unlike the two steps involved in AKC-Composite, the final calibrated weights for MR-Composite are obtained directly.

4.2 Explicit Formulas for MR-Composite.

Suppose x denotes all the auxiliary variables and is partitioned into two subsets, $x^{(1)}$ and $x^{(2)}$, the first subset corresponds to the "old" auxiliary variables normally used in GR and the second subset corresponds to the "new" auxiliary variables under MR representing key study variables from t and t'. The MR-Composite can then be expressed as

$$\hat{\theta}_{yt}^{MR} = \hat{\theta}_{yt}^{HT} + (\hat{\beta}_t^{MR})' \, (\tilde{\theta}_{xt} - \hat{\theta}_{xt}^{HT})$$

where $\tilde{\theta}_{xt}$ includes both fixed (for old x) and random (for new x) population totals, i.e., $\tilde{\theta}_{xt} = (\theta_{xt(1)}, \hat{\theta}_{xt(2)}^{MR})$, and $\hat{\beta}_t^{MR} = (y_t' \Lambda_t X_t)\,(X_t' \Lambda_t X_t)^{-1}$, where $\Lambda_t = \mathrm{diag}(h_{kt})$, $X_t = (X_t^{(1)}, X_t^{(2)})$.

Alternatively, $\hat{\theta}_{yt}^{MR}$ can be expressed as $\hat{\theta}_{yt}^{MR} = \hat{\theta}_{yt}^{GR} + (\hat{\beta}_{t2.1}^{MR})' \, (\hat{\theta}_{xt(2)} - \hat{\theta}_{xt(2)}^{GR})$, where $\hat{\beta}_{t2.1}^{MR}$ are the partial regression coefficients $[\, y_t' \, \Lambda_t (I - PX_t^{(1)}) X_t^{(2)}] \times [X_t^{(2)\prime} \, \Lambda_t (I - PX_t^{(1)}) X_t^{(2)}]^{-1}$, and $PX_t^{(1)}$ is the projection operator $X_t^{(1)} \, (X_t^{(1)\prime} \, \Lambda_t X_t^{(1)})^{-1} \, X_t^{(1)\prime} \, \Lambda_t$. This alternative expression of $\hat{\theta}_{yt}^{MR}$ is useful for direct comparison with the form of AK-Composite.

4.3 Variance Estimation for MR-Composite.

It follows from the Taylor linearization variance formula for the GEF estimator that for the estimating equation $G_t' \, \Gamma_t^{-1} \, g_t = 0$, the asymptotic variance of $\hat{\theta}_{yt} - \theta_{yt}$ has the form, $B' \, G_t' \Gamma_t^{-1} \hat{V}(g_t) \Gamma_t^{-1} \, G_t \, (B^{-1})'$, where $\hat{V}(g_t)$ is a consistent estimate of the true covariance of g_t, and $B = G_t' \, \Gamma_t^{-1} \, G_t$.

4.4 Theoretical Comparison with Existing Methods.

It is easily seen that the limitations (a) to (d) of the AK-Composite, which were not overcome by AKC-Composite, are indeed overcome by MR-Composite. The price paid for this is the relaxation of the optimality criterion to that of suboptimality via the use of the working covariance matrix. The resulting loss in efficiency may be more than offset by the possible gain due to the addition of the multivariate aspect. With MR-Composite, all the estimates for linear functions (such as aggregates) of the selected study variables are indeed MR-Composite, and in the case of linearly dependent variables, the estimates are invariant with respect to the choice of the left over variable. These properties follow directly by viewing MR as a multivariate regression estimator.

5. EMPIRICAL STUDY

Using monthly LFS data for Ontario for the year 1993, the MR-Composite was compared with AK-Composite for variables selected for composite estimation for both level and month-to-month changes. For a few other variables, it was compared with AKC-Composite. The evaluation measure used was gain in efficiency relative to GR averaged over the year; it was calculated in percent as a ratio of variances of GR and the alternative estimate minus one times 100.

LFS currently uses GR. The population totals for auxiliary variables used for GR correspond to 30 (15x2 age-sex groups), 10 census metropolitan areas, 11 economic regions (they aggregate to the total population count), 8 urban centres, and the 6 rotation groups. For MR-Composite, only a subset of auxiliary variables corresponding to the key study variables for current and previous occasions was used. In particular, only three additional variables E, U, and N for the previous month were used. For these variables, macro-level information in terms of the full sample estimates from the previous month and micro-level information for the matched subsample were extracted. The full sample estimates were adjusted for change in the population from the previous to the current month. For simplicity, we used a multiplicative adjustment as $\hat{\theta}_{y't'}^{MR}(P/P')$ where P is the 15+ population for month t and P' for month t', instead of the additive adjustment $\hat{\theta}_{y't'}^{MR} + (\hat{\theta}_{y'tm}^{GR} - \hat{\theta}_{y't'm'}^{GR})$ as mentioned in Section 4.

For the AK-Composite, (A,K) values for E and U were selected as (.2,.4) and (.55,.55) respectively. These are averages of optimal values obtained by grid search for the period Aug-Dec,'92. The starting point for all recursive estimates for 1993 was taken as July, '92 allowing for a six month break-in period. Variances were computed using the jackknife resampling method since this is currently in use for LFS. Pseudo-replicates for the previous month full sample estimate were used to account for the extra variability due to random population totals in composite estimation.

Tables 1 and 2 show average gains in efficiency for AK-Composite and MR-Composite relative to GR. AKC-Composite was used for variables not used as auxiliary. These are labour force status by sex (M or F), by full-time (FT) or part-time (PT), and the two rates, unemployment rate and participation rate defined respectively as U/(E+U) and (E+U)/P. It is seen that MR generally outperforms AKC for both level and change estimates.

6. CONCLUDING REMARKS

MR-Composite uses cross-fertilization of ideas from different areas of statistics, e.g., the theory of zero functions, MEF (or GEF) and its relation to regression, the idea of partial overlap in the d-based treatment of estimation from time series of survey estimates, and the methodology of GR and WCS in survey sampling. The GEF theory is seen to give a fresh perspective to the problem which is known to have a long history. Use of GEF, as an alternative to the well known model-assisted framework in survey sampling, provided the necessary ingredients for making departures from the tradition. Finally, in view of the promising results of the empirical study, MR-Composite is currently being considered for implementation in the production of LFS-based estimates.

REFERENCES

Binder, D.A., and Hidiroglou, M.A. (1988). Sampling in time. *Handbook of Statistics*, Vol. 6 (P.R. Krishnaiah and C.R. Rao, eds.), 187-211.

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

Godambe, V.P. and Thompson, M.E. (1989). An extension of quasi-likelihood estimation (with discussion). *Journal of Statistical Planning and Inference*, 22, 137-72.

Gurney, M. and Daly, J.F. (1965). A multivariate approach to estimation in periodic sample surveys. *Proc. Amer. Statist. Assoc., Sec. Sur. Res. Meth.*, 247-257.

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

Lent, J., Miller, S. and Cantwell, P. (1994). Composite weights for the current population surveys. ASA Proc. Surv. Res. Meth. Sec.

Rao, J.N.K. and Graham, J.E. (1964). Rotation designs for sampling on repeated occasions. *JASA*, 50, 492-509.

Singh, A.C. (1994). Sampling design-based estimating functions for finite population totals. Invited paper, Abstracts of the Annual Meeting of the Statistical Society of Canada, Banff, Alberta, May 8-11; 48.

Table 1: Average Gain in Efficiency for Levels

Variable	MR	AK (or AKC)
E	19.47	7.74
U	15.93	10.53
N	11.82	4.04
E,M	12.04	4.96
E,F	8.96	4.34
U,M	9.07	6.88
U,F	5.22	3.88
N,M	5.71	0.83
N,F	7.46	3.79
E, FT	12.93	4.56
E, PT	2.22	1.58
U-RATE	17.61	11.00
P-RATE	11.61	4.01

Table 2: Average Gain in Efficiency for Change

Variable	MR	AK (or AKC)
E	16.56	12.85
U	24.39	12.42
N	4.88	2.89
E,M	11.84	8.51
E,F	1.95	3.96
U,M	12.13	6.82
U,F	8.12	5.01
N,M	2.63	0.76
N,F	0.55	2.06
E, FT	7.4	5.78
E, PT	2.08	2.18