COMBINING INFORMATION IN SURVEY SAMPLING BY MODIFIED REGRESSION

Avinash C. Singh, Statistics Canada and Carleton University
Survey and Analysis Methods Development, Statistics Canada, 16-A, R.H. Coats, Ottawa, Ontario K1A 0T6

ABSTRACT

For complex surveys, a design-based methodology termed modified regression is proposed. It uses the idea of finite population (semiparametric) modelling with a working covariance structure within the (generalized) zero function framework. Also it encompasses the familiar design-based methodology of generalized regression which uses the idea of superpopulation modelling within the model-assisted framework. It is shown that the method of modified regression provides a unified approach to estimation in several problems of combining information in survey sampling such as those arising from two occasions, two frames, two phases, small areas and outlier-prone domains. The problem of estimating nonlinear parameters, such as the median, is also covered by the proposed method. For estimating finite population parameters in the context of survey sampling, the proposed method of modified regression can be viewed as an analogue of the method of (generalized) zero functions for estimating infinite population parameters in the context of classical statistics.

Key Words: Finite population model; Predictor zero functions; Weight calibration; Working covariance.

1. INTRODUCTION

We consider estimation of finite population parameters under the design-based approach, i.e., values of the study and auxiliary variables attached to units in the finite population are assumed to be fixed, and the only source of randomization of the sample observations is due to a given probability sampling design $p(s)$. The parameters of interest could be linear (such as total) or nonlinear (such as median); we will say that a parameter is linear if it can be expressed as a population total of values of a function of the study variable. The problem considered is the estimation of a parameter when in addition to the sample data $y_1, ..., y_n$, correlated auxiliary information is available. The auxiliary information could be in several forms or combination of them: (i) known population total for a correlated variable x (or an estimate of the total based on a larger sample of x-values), (ii) additional (cross-sectional) sample of y-values which may represent a population partially overlapping with the target population, (iii) past sample from a partially overlapping longitudinal survey, and (iv) prior information about the parameter of interest.

We propose a method termed modified regression (MR) which was first introduced by Singh (1994) in a somewhat different manner, and was inspired by the contributions of Fuller (1975), Särndal (1980), and Rao and Scott (1981) in survey sampling for estimating finite population parameters, and C.R. Rao (1968), Henderson (1975), Nelder and Wedderburn (1972), Wedderburn (1974), Liang and Zeger (1986), and Godambe and Thompson (1989) in classical statistics for estimating infinite population parameters. In the classical statistics, one could use a common principle for various problems of estimation involving combining correlated information. This has to do with the use of linear zero functions of Rao (1968) (which gives a convenient alternative characterization of Gauss-Markov) to get optimal (BLUE) estimates of linear (semiparametric) model parameters, and more generally, use of maximum quasi-likelihood (Wedderburn, 1974), and optimal estimating functions (Godambe and Thompson, 1989) for nonlinear model parameters. When some parameters are random with known prior distribution (up to second moments), then optimal (BLUP) estimates (Henderson, 1975, see also Robinson, 1991) for linear model parameters can be obtained, and for nonlinear model parameters, a generalization of estimating functions given by generalized zero (or predicting) functions (Singh, 1995) can be used. When the covariance structure is difficult to specify, then the use of a suitable working covariance matrix (as in the GEE-generalized estimating equations approach of Liang and Zeger, 1986) can be used which gives rise to suboptimal but consistent estimates in general. Thus, the method of generalized zero functions for semiparametric model parameters (which gives rise to above methods as special cases) provides the underlying common principle.

The above general principle of estimation by generalized zero functions used in classical statistics can also be used in survey sampling provided a suitable finite population (semiparametric) model can be defined. Although it may not be possible to define such a model at the unit level (in view of the noninformativeness of the design-based likelihood with labelled units, Godambe 1966), it may be possible to define it at the aggregate level. Suppose, the sample data is condensed into elementary estimates corresponding to subsamples (or domains) of interest using Horvitz-Thompson estimation. Now, treating these elementary estimates as working sufficient statistics (i.e., as building blocks), and assuming that their design-based covariance matrix can be specified, one can use optimal regression via linear zero functions (or generalized zero functions for the general case of nonlinear or random model parameters) to get estimates for finite population parameters of interest. However, this solution is not practical except for simple designs because the covariance structure is difficult to specify for complex designs. An obvious alternative is to use the idea of working covariance matrix (as in GEE) and define suboptimal estimates. This is the route followed by the proposed MR method. The MR methodology provides a general principle for estimation when combining information from two phase sampling, two frames, two time points, as well as when prior information is available.The parameter of interest may be linear or nonlinear.The option of using prior information by treating the finite-population parameter as random is particularly useful in dealing with the problem of small area and outlier-prone domain estimation, in which case MR leads to biased but stable estimates under design-based framework.(Under the design-cum-prior distribution, the estimator will be (approximately) unbiased). Note that estimation of design-based MSE would, however, be difficult in this case due to design bias.

For problems of combining information, the existing design-based methods in survey sampling often employ the concept of superpopulation modelling. For example, for the problem of combining auxiliary information from known population totals for correlated variables, both Fuller (1975) and Särndal (1980) used a superpopulation model to motivate regression-type estimators (known as model-assisted); Särndal's generalized regression (GR) estimator is more general and encompasses Fuller's estimator. For the problem of smoothing (categorical) finite population parameter estimates under a hypothesis (which can be viewed as a problem of combining information in the form of parameter restriction imposed by the hypothesis), Rao and Scott (1981) used the method of pseudo maximum likelihood (PML); this can also be motivated from a superpopulation model. However, for various problems of combining information mentioned earlier, it may be difficult to justify a suitable superpopulation model required for a model-assisted estimator. Besides, if the design-based properties are of interest, the superpopulation model remains somewhat passive. The MR approach, on the other hand, avoids the need of superpopulation modelling by using the idea of finite population modelling with a working covariance structure, and can recover the existing model-assisted methods as special cases.

Based on the above discussion, it follows that use of MR in estimating finite population parameters provides a common principle which is analogous to the use of generalized zero functions for infinite population parameters. However, there is an important aspect in which the two differ. Since the probability sampling mechanism depends on known design variables and not on the study variable y, it is possible to develop an estimation method which is not y-specific, i.e., which is applicable to all study variables. This feature is very appealing in practice and is known as weight calibration in sampling (Deville and Särndal, 1992). For infinite population parameters, on the other hand, the estimation is y-specific because generally the model specification depends intrinsically on the variable y. For each MR-estimator considered, we provide a representation in terms of expansion estimates using calibrated weights which will be common for all study variables. Note that for this purpose a common set of auxiliary variables is needed for all study variables. However, variance of the estimator for a given study variable will of course depend on the correlation between the study and the chosen auxiliary variables.

The organization of this paper is as follows. Section 2 provides a motivation of the proposed MR method while Section 3 contains its description. Illustrative examples of its application are given in Section 4. Finally, Section 5 contains concluding remarks.

2. MOTIVATION OF MODIFIED REGRESSION

To motivate MR, we will first show how the usual generalized regression (GR) estimator, developed by Särndal (1980) in a model-assisted framework, can be obtained by finite population (and not superpopulation) modelling. In GR, the extra information is in the form of known population totals $\boldsymbol{\theta}_x$ of p-vector of auxiliary variables $\boldsymbol{x}$. This gives rise to p predictor zero functions $\hat{\boldsymbol{\theta}}_x^{HT} - \boldsymbol{\theta}_x$ where the p-vector $\hat{\boldsymbol{\theta}}_x^{HT} = \boldsymbol{X}' \boldsymbol{W} \boldsymbol{1}_n$ is the Horvitz-Thompson estimator of $\boldsymbol{\theta}_x$, $\boldsymbol{W} = diag(\boldsymbol{h}_{n\times 1})$, $\boldsymbol{X}$ is a $n \times p$ matrix of observed $\boldsymbol{x}$-values, and $\boldsymbol{h}$ is the n-vector of inverse inclusion probabilities. Now, the GR-estimator of the total θ_y of the study variable y is given by

$$\hat{\theta}_y^{GR} = \hat{\theta}_y^{HT} - \boldsymbol{y}' \boldsymbol{W}\boldsymbol{X}(\boldsymbol{X}' \boldsymbol{W}\boldsymbol{X})^{-1} (\hat{\boldsymbol{\theta}}_x^{HT} - \boldsymbol{\theta}_x) \quad (2.1)$$

The above estimator can be obtained using the method of linear zero functions, i.e., as a residual after regressing $\hat{\theta}_y^{HT}$ on the p-predictors under a working covariance $\boldsymbol{\Gamma}_g$ (see 2.2b) of the $(p+1)$-vector $\boldsymbol{g} = (\hat{\theta}_y^{HT} - \theta_y, (\hat{\boldsymbol{\theta}}_x^{HT} - \boldsymbol{\theta}_x)')'$. The underlying finite population model is a semiparametric common mean model given by

$$\begin{bmatrix} \hat{\theta}_y^{HT} \\ \hat{\theta}_y^{HT} + (\hat{\theta}_{x1}^{HT} - \theta_{x1}) \\ \cdot \\ \cdot \\ \cdot \\ \hat{\theta}_y^{HT} + (\hat{\theta}_{xp}^{HT} - \theta_{xp}) \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ \cdot \\ \cdot \\ \cdot \\ 1 \end{bmatrix} \theta_y + \epsilon \quad (2.2a)$$

where $\epsilon_{(p+1)\times 1} \sim (\boldsymbol{0}, \boldsymbol{\Gamma}_\epsilon)$, $\boldsymbol{\Gamma}_\epsilon = \boldsymbol{C}\boldsymbol{\Gamma}_g\boldsymbol{C}'$,

$$\boldsymbol{C} = \begin{pmatrix} 1 & \boldsymbol{0}'_{1\times p} \\ \boldsymbol{1}_{p\times 1} & \boldsymbol{I}_{p\times p} \end{pmatrix}, \quad \boldsymbol{\Gamma}_g = \begin{pmatrix} \boldsymbol{y}'\boldsymbol{W}\boldsymbol{y} & \boldsymbol{y}'\boldsymbol{W}\boldsymbol{X} \\ \boldsymbol{X}'\boldsymbol{W}\boldsymbol{y} & \boldsymbol{X}'\boldsymbol{W}\boldsymbol{X} \end{pmatrix} \quad (2.2b)$$

The matrix $\boldsymbol{\Gamma}_g$ is of course nnd because it can be expressed as $(y,X)' W(y,X)$, and $\boldsymbol{W}$ is nnd. However at first sight it may seem unusual to treat $\boldsymbol{\Gamma}_g$ as a working covariance because its entries do not involve deviations from the mean. We will show that under suitable conditions, the above choice of $\mathbf{\Gamma}_g$ gives rise to optimal regression estimate for simple random samples, but suboptimal estimates in general. Thus there is some justification in choosing $\mathbf{\Gamma}_g$. This can be seen as follows. The conditions are: (i) the population size N is known and $\boldsymbol{h}'\mathbf{1}_n = N$; a condition often met in practice by using $\hat{\theta}_y^{HHT}$ (the Hajek-modification of the HT-estimator) defined as $\boldsymbol{X}'\boldsymbol{W}_H\mathbf{1}_n$ where $\boldsymbol{W}_H = diag\{h_k N(\sum h_k)^{-1}: k \in s\}$, (ii) the counting variable is used as one of $\boldsymbol{x}$. Equivalently, this means that $\mathbf{1}_n$ is in the column space of $\boldsymbol{X}$. We will assume without loss of generality that the first variable x_1 is 1 so that $\theta_{x1} = N$. Now, denoting by $\boldsymbol{x}_{(1)}$ the x-vector without the first element, the estimator $\hat{\theta}_y^{GR}$ obtained under the working covariance matrix $\mathbf{\Gamma}_g$ is equivalent to the estimator obtained by optimal regression of $\hat{\theta}_y^{HHT}$ under simple random sampling on $(p-1)$ predictor zero functions corresponding to $\boldsymbol{x}_{(1)}$, see (2.6) below.

More specifically, under the condition (i), the GR-estimator has the simple projection form which in turn gives the following results.

$$\hat{\theta}_y^{GR} = \boldsymbol{y}'\boldsymbol{W}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{\theta}_x \quad (2.3a)$$
$$= \boldsymbol{y}'\boldsymbol{W}_H\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{W}_H\boldsymbol{X})^{-1}\boldsymbol{\theta}_x \quad (2.3b)$$
$$= \hat{\theta}_y^{HHT} - \boldsymbol{y}'\boldsymbol{W}_H\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{W}_H\boldsymbol{X})^{-1}(\hat{\boldsymbol{\theta}}_x^{HHT} - \boldsymbol{\theta}_x) \quad (2.3c)$$
$$= \hat{\theta}_y^{HHT} - \boldsymbol{y}'\boldsymbol{W}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{W}\boldsymbol{X})^{-1}(\hat{\boldsymbol{\theta}}_x^{HHT} - \boldsymbol{\theta}_x) \quad (2.3d)$$

Note that although the first element of $\boldsymbol{\theta}_x - \hat{\boldsymbol{\theta}}_x^{HHT}$ is zero under the condition (ii), its contribution in the working covariance $\mathbf{\Gamma}_g$ is not zero because $\mathbf{\Gamma}_g$ does not involve deviations from the mean. Now using the inverse partitioned matrix formula for the matrix

$$\boldsymbol{X}'\boldsymbol{W}\boldsymbol{X} = \begin{vmatrix} \boldsymbol{x}_1'\boldsymbol{W}\boldsymbol{x}_1 & \boldsymbol{x}_1'\boldsymbol{W}\boldsymbol{X}_{(1)} \\ \boldsymbol{X}_{(1)}'\boldsymbol{W}\boldsymbol{x}_1 & \boldsymbol{X}_{(1)}'\boldsymbol{W}\boldsymbol{X}_{(1)} \end{vmatrix}, \quad (2.4)$$

where $\boldsymbol{X}_{(1)}$ denotes the matrix $\boldsymbol{X}$ without first row and column, we obtain from (2.3d)

$$\hat{\boldsymbol{\theta}}_y^{GR} = \hat{\theta}_y^{HHT} - (\boldsymbol{y}'\boldsymbol{W}\boldsymbol{X}_{c(1)})(\boldsymbol{X}_{c(1)}'\boldsymbol{W}\boldsymbol{X}_{c(1)})^{-1}(\hat{\boldsymbol{\theta}}_{x(1)}^{HHT} - \boldsymbol{\theta}_{x(1)}). \quad (2.5)$$

Here $\boldsymbol{X}_{c(1)}$ denotes the centered $\boldsymbol{X}_{(1)}$ matrix, i.e., from column i, the sample mean $\bar{x}_i (= \sum_{k\in s} x_{ik} h_k (\sum h_k)^{-1})$ is subtracted. Thus, $\boldsymbol{X}_{c(1)} = \boldsymbol{X}_{(1)} - \mathbf{1}_n \otimes \bar{\boldsymbol{x}}_{(1)}$. We can write (2.5) as

$$\hat{\theta}_y^{GR} = N\bar{y} - [\textstyle\sum_{k=1}^n y_k (\boldsymbol{x}_{k(1)} - \bar{\boldsymbol{x}}_{(1)}) h_k]$$
$$\times [\textstyle\sum_{k=1}^n (\boldsymbol{x}_{k(1)} - \bar{\boldsymbol{x}}_{(1)})(\boldsymbol{x}_{k(1)} - \bar{\boldsymbol{x}}_{(1)})' h_k]^{-1} (N\bar{\boldsymbol{x}}_{(1)} - \boldsymbol{\theta}_{x(1)}) \quad (2.6)$$

Note that the sampling weights h_k are included in the simple random sample variance-covariance expressions. Thus (2.6) coincides with the optimal regression estimator for simple random samples.

The formulation of the finite population model (2.2) in terms of zero functions defines MR for the special case of GR. As will be seen in the next section, MR gives GR-type solutions to many different problems. Now there are several ways in which MR can generalize GR. For example, MR allows for other zero functions such as those given by the difference of two estimates. The zero functions can also involve nonlinear parameters. Also MR gives rise to a family of estimates by choosing different working covariance matrices. Moreover, prior information about the parameter can be incorporated by using additional zero functions. Finally, like GR, MR-estimates for any y can be represented as an expansion estimator, $\boldsymbol{y}'\boldsymbol{w}$ where $\boldsymbol{w}$ represents the n-vector of calibrated weights common for all y-variables. In particular, for GR, the calibrated weights are given by

$$\boldsymbol{w}^{GR} = \boldsymbol{h} - \boldsymbol{W}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{W}\boldsymbol{X})^{-1}(\hat{\boldsymbol{\theta}}_x^{HT} - \boldsymbol{\theta}_x) \quad (2.6)$$

The variance of the MR-estimator can be estimated in general by a sandwich-type estimator as in the case of estimating equations (see e.g. Binder, 1983). In particular, for GR, the usual variance formula in terms of residuals can be expressed alternatively as follows. First note that $\hat{\theta}_y^{GR}$ is the solution of the following estimating equation

$$\boldsymbol{G}'\mathbf{\Gamma}_g^{-1}\boldsymbol{g} = 0 \quad (2.7)$$

where $\boldsymbol{G} = (1, 0, ..., 0)'$, $\boldsymbol{g} = (\hat{\theta}_y^{HT} - \theta_y, (\hat{\boldsymbol{\theta}}_x^{HT} - \boldsymbol{\theta}_x)')'$.

Therefore,

$$\hat{V}(\hat{\theta}_y^{GR}) = B^{-1}(G'\Gamma_g^{-1}\hat{V}_g\Gamma_g^{-1}G)(B^{-1})', \quad (2.8)$$

where $B = G'\Gamma_g^{-1}G$, and $\hat{V}_g$ is an estimate of the true covariance matrix of the vector $\boldsymbol{g}$ of HT-estimates using standard formulas in sampling. Under the asymptotic framework of Isaki and Fuller (1982), the estimator $\hat{\theta}_y^{GR}$ is asymptotically normal with mean θ_y and variance $\hat{V}(\hat{\theta}_y^{GR})$ as $n, N \to \infty$. This follows from the CLT of $\boldsymbol{G}'\mathbf{\Gamma}_g^{-1}\boldsymbol{g}$. This implies, in particular, that $\hat{\theta}_y^{GR}$ is asymptotically design consistent, i.e.

$$N^{-1} \mid \hat{\theta}_y^{GR} - \theta_y \mid = O_p(n^{-1/2}), \text{ since } \hat{V}(\hat{\theta}_y^{GR}) = O_p(n^{-1}N^2).$$

3. METHOD OF MODIFIED REGRESSION

In GR, we worked with the elementary unbiased estimates of finite population parameters for study and auxiliary variables. However, for defining MR, it would be useful to work with zero functions which are elementary unbiased estimates of zero; here the function may depend in general on both the variable and the parameter. This makes it convenient to deal with nonlinear parameters. The function $g_\phi(y, \theta)$ will be used to denote the (parameter-dependent) elementary zero function involving the study variable and the parameter of interest θ. Thus,

$$g_\phi(y,\theta) = \textstyle\sum_{k\in s} \phi(y_k,\theta)h_k - \sum_{k\in U}\phi(y_k,\theta) \quad (3.1a)$$
$$= \textstyle\sum_{k\in U}(I_{\{k\in s\}} - h_k^{-1})\phi(y_k,\theta)h_k \quad (3.1b)$$

If $\phi(y_k, \theta)$ does not depend on θ, then the finite population parameter θ implied by (3.1) is the population total $\theta_\phi = \sum_{k\in U}\phi(y_k)$. (If $\phi(y_k) = y_k$, then θ_ϕ is simply θ_y). In this case, we will say that the parameter is linear because the parameter θ_ϕ can be expressed as a population total of values of a function of y_k. However, if $\phi(y_k, \theta)$ does depend on θ, then the parameter is nonlinear in population values, in which case θ is defined by solving $\sum_{k\in U}\phi(y_k, \theta) = c$ for a known value of c. For example, if θ_y is the finite population median, then $\phi(y_k,\tilde{\theta}_y) = I\{y_k \le \tilde{\theta}_y\}$, and $c = N/2$.

Similarly, the functions $g_\psi(x)$ will be used to denote the (parameter-free) elementary zero functions involving the auxiliary variables x. For example, g_ψ for GR can be expressed as, for $i=1, ..., p$, and $\psi(x_{ik}) = x_{ik}$,

$$g_\psi(x_i) = \sum_s \psi(x_{ik})h_k - \theta_\psi \qquad (3.2a)$$
$$= \sum_{k\in U}(I_{\{k\in s\}} - h_k^{-1})\,\psi(x_{ik})h_k \qquad (3.2b)$$

The function g_ψ is parameter-free because $\theta_\psi = \sum_{k\in U}\psi(x_{ik})$ is assumed to be known. Now, for other types of predictor zero functions such as the difference of two estimates obtained from the same sample, $\psi(x_{ik})$ can be suitably defined as a function of x_{ik} such that $\sum_{k\in U}\psi(x_{ik})=0$.

We distinguish between two scenarios depending on whether prior information is available for θ or not.

3.1 No Prior Information About θ

Case I (Linear Case)

Here, $\theta = \sum_{k\in U}\phi(y_k)$. For combining information about θ in terms of $(p+1)$ zero functions $g_\phi(y,\theta)$ and $g_\psi(x)_{px1}$, it follows from the method of zero functions that the optimal estimator is given by the regression estimator for the common mean model similar to (2.2) provided the true covariance matrix V_g of the vector of zero functions $g = (g_\phi, g_\psi)'$ is (approximately) known. It is obtained as the solution of the following estimating equation,

$$G'\,V_g^{-1}\,g = 0, \qquad (3.2)$$

where $G' = (1, 0, ..., 0)_{1\times(p+1)}$. However, for complex designs, it is in general difficult to find a stable estimate of V_g due to insufficient degrees of freedom (cf: Rao, 1994), and therefore a working covariance matrix Γ_g can be used instead. This defines MR for the linear case. In particular, if Γ_g is chosen in a manner similar to that for GR, i.e.,

$$\Gamma_g = \begin{bmatrix} \phi'\,W\phi & \phi'\,W\Psi \\ \Psi'\,W\phi & \Psi'\,W\Psi \end{bmatrix}, \qquad (3.3)$$

where ϕ is the $n\times 1$ vector of ϕ-values, Ψ is the $n\times p$ matrix of ψ-values ($k=1$ *to* n, $i=1$ *to* p), then the estimator $\hat{\theta}^{MR}$ and calibrated weights w^{MR} are given by

$$\hat{\theta}^{MR} = \sum_{k\in s}\phi(y_k)\,w_k^{MR} = \phi'w^{MR} \qquad (3.4a)$$
$$w^{MR} = h - W\Psi(\Psi'\,W\Psi)^{-1}\,g_\psi. \qquad (3.4b)$$

In various applications, the following more general form of the calibrated weights for MR will be used.

$$w_*^{MR} = h_* - W_*\,\Psi_*(\Psi_*'\,W_*\,\Psi_*)^{-1}\,g_{\psi *}. \qquad (3.5)$$

This entails replacing (i) predictor functions g_ψ by $g_{\psi *}$ (and the corresponding covariate matrix Ψ by Ψ_*); this feature allows for transforming predictor functions in order to have certain desirable properties in MR, see e.g., composite estimation (Section 4.1) and two-phase estimation (Section 4.3), (ii) the weight matrix W by W_*; here unlike W, W_* is not necessarily $diag(h_*)$, see e.g., dual frame estimation in Section 4.2, and (iii) the initial weight vector h by h_*; here h_*, unlike h, need not be the usual vector of inverse inclusion probabilities, see e.g., two-phase estimation in Section 4.3 where h_* denotes the inverse of conditional inclusion probabilities at the second phase given the first phase sample, and small area estimation to allow biased (but asymptotically design consistent) estimation in Section 4.5. The properties of the underlying finite population model should be kept in mind to suitably choose the predictors and the working covariance which in turn define Ψ_*, h_*, W_* for the calibration form of MR. A direct application of calibration estimation for a given distance function may not be justifiable.

Now as before for GR, $\hat{V}(\hat{\theta}^{MR})$ can be obtained in a sandwich form (2.8), and an approximate confidence interval for θ can be constructed from the asymptotic normality of $\hat{\theta}^{MR} - \theta$ with mean 0 and variance $\hat{V}(\hat{\theta}^{MR})$.

Case II (Nonlinear Case)

Here θ is defined as a solution of $\sum_{k\in U}\phi(y_k,\theta) = c$. Now, in general, it may not be possible to write a common mean model as in the linear case from the zero functions $g_\phi(y,\theta)$ and $g_\psi(x)$ because y and θ may not separate in the expression $g_\phi(y,\theta)$. However, the method of estimating functions of Godambe and Thompson (1989), which is more general than the method of linear zero functions, can be adapted for the finite population parameters under certain conditions. Here the finite population semiparametric model can be expressed as

$$g \sim (0, V_g). \qquad (3.6)$$

Suppose for the class of estimating functions $A'g$ defined by $n\times 1$ transformation vectors A, we have the following asymptotic representation as n, $N\to\infty$,

$$A'\,V_g^{-1}g = A'\,V_g^{-1}G(\theta-\hat{\theta}) + o_p(n^{-1/2}N), \qquad (3.7)$$

where $G = (f(\theta), 0, ..., 0)'$ for a suitable function $f(\theta)$. Then the optimal choice of A is given by G. Note that (3.7) is not a Taylor expansion where G represents the vector of derivatives of g with respect to θ, because in the context of finite population parameters, these may not be meaningful. However, these can be viewed as pseudo-derivatives. Thus, under the condition (3.7), the optimal estimating function has the same form as (3.2) of the linear case.

Using the working covariance matrix Γ_g of (3.3), the MR-estimator for the nonlinear case is obtained by solving

$$\phi'(y,\theta)\,w^{MR} = c, \qquad (3.8)$$

where w^{MR} has the same form (3.5) as in the linear case, and is independent of θ. Notice that the estimating equation (3.8) and hence the estimator $\hat{\theta}^{MR}$ does not depend on the pseudo-derivative $f(\theta)$. However, its variance, $V(\hat{\theta}^{MR})$ would, of course, depend on $f(\theta)$. It is interesting to note that MR justifies a natural and commonly used estimator of θ. In the absence of auxiliary information, θ is estimated by solving $\sum_{k\in s}\phi_k(y_k,\theta)h_k = c$ while in the presence of auxiliary information, h_k is replaced by w_k^{MR} and θ is estimated by solving $\sum_{k\in s}\phi_k(y_k,\theta)\,w_k^{MR} = c$. From a practical point of view, the resulting estimator of θ is convenient because it uses the same set of calibrated weights as in the linear case. However, choice of a suitable set of predictor variables would invariably depend on whether the parameter is linear or not, see also Section 4.4.

3.2 Prior Information About θ

We will consider only the linear parameter case. If the estimator $\hat{\theta}^{MR}$ is unstable due to insufficient sample size as in the case of small area or outlier-prone domain estimation, it may be worthwhile to reduce instability by using prior information at the cost of introducing (marginal) design bias. Using the method of generalized zero functions, it is easy to combine the additional prior information. The prior information about θ can be expressed in the form of a zero function (under the prior distribution), $g_\theta = \theta^* - \theta$, where θ^* is the prior estimate of θ. Since the zero functions $\boldsymbol{g} = (g_\phi, g_\psi^{\prime})^{\prime}$ and g_θ are uncorrelated, it is easy to see that the optimal (under design-cum-prior distribution) estimator, $\hat{\theta}^{OPT}$ will be a convex linear combination of $\hat{\theta}^{MR}$ and θ^*, and is given by

$$\hat{\theta}^{OPT} = \lambda\hat{\theta}^{MR} + (1-\lambda)\theta^*, \qquad (3.9a)$$

where $$\lambda = V(g_\theta)[V(g_\theta) + V(\hat{\theta}_{MR})]^{-1} \qquad (3.9b)$$

Using a working covariance for $\boldsymbol{g}$ to compute λ, the resulting estimator, $\hat{\theta}^{MR*}$ (say), would be suboptimal. Under only design-based framework, the estimator $\hat{\theta}^{MR*}$ is biased, but more stable (i.e., has less variance) than $\hat{\theta}^{MR}$, but its MSE is, in general, difficult to estimate. However, biased estimates can be quite useful in practice if it can be ensured that bias (relative to standard error) is not too high.

In practice, some working assumptions may be used to choose the prior estimate θ^* and the shrinkage factor λ. For example, if θ^* is set equal to $\hat{\boldsymbol{\beta}}^{\prime}\boldsymbol{\theta}_\psi$, where $\hat{\boldsymbol{\beta}}^{\prime} = \boldsymbol{\phi}^{\prime}\boldsymbol{\Gamma}\boldsymbol{\Psi}(\boldsymbol{\Psi}^{\prime}\boldsymbol{\Gamma}\boldsymbol{\Psi})^{-1}$, and $\boldsymbol{\theta}_\psi = \sum_{k\in U}\psi(\boldsymbol{x})$, then $\hat{\theta}^{MR*}$ assumes the form of a calibration estimator, i.e.,

$$\hat{\theta}^{MR} = \boldsymbol{\phi}^{\prime}\boldsymbol{w}^{MR*} \qquad (3.10a)$$

where $$\boldsymbol{w}^{MR*} = \lambda\boldsymbol{h} - \boldsymbol{\Gamma}\boldsymbol{\Psi}(\boldsymbol{\Psi}^{\prime}\boldsymbol{\Gamma}\boldsymbol{\Psi})^{-1}\boldsymbol{g}_{\psi*}, \qquad (3.10b)$$

$$\boldsymbol{g}_{\psi*} = \lambda\sum_{k\in s}\psi(\boldsymbol{x}_k)h_k - \boldsymbol{\theta}_\psi = \sum_{k\in U}(\lambda I_{\{k\in s\}} - h_k^{-1})\psi(\boldsymbol{x}_k)h_k$$

Thus, with a working value of λ, the estimator $\hat{\theta}^{MR*}$ can be easily calculated from a simple modification of the usual regression weight calibration. In practice, $\hat{\boldsymbol{\beta}}$ itself can be made more stable by borrowing ϕ and ψ values suitably (see Section 4.5 and 4.6). This modification helps to make $\hat{\theta}^{MR}$ approximately unbiased, but the problem of instability in $\hat{\theta}^{MR}$ still persists. Also, in practice, the value of λ is chosen such that it tends to 1 as $n \to \infty$. Therefore, for large samples, $\hat{\theta}^{MR*}$ will be asymptotically design consistent.

4. EXAMPLES OF MODIFIED REGRESSION

4.1 Composite Estimation by MR

With rotating panel surveys, usual estimates of level and change based on only cross-sectional data can be made more efficient by incorporating in the estimates correlated information from another time point due to overlapping samples. The traditional estimator is the AK-estimator of Gurney and Daly (1965). Let t, $t^{\prime}$ denote the current and previous time points and let y, m denote respectively the study variable at t, and the matched (backward with $t^{\prime}$-sample) subsample of t-sample. Similarly, $y^{\prime}, m^{\prime}$ are defined where $m^{\prime}$ now denotes the matched (forward with t-sample) subsample of $t^{\prime}$-sample. The AK-composite estimator, $\hat{\theta}_{yt}^{AK}$, of the total θ_{yt} for the study variable y at time t is given by $\hat{\theta}_{yt}^{AK} =$

$$\hat{\theta}_{yt}^{GR} + K[\hat{\theta}_{y^{\prime}t^{\prime}}^{AK} + (\hat{\theta}_{ytm}^{GR} - \theta_{y^{\prime}t^{\prime}m^{\prime}}^{GR}) - \hat{\theta}_{yt}^{GR}] + A(\hat{\theta}_{yt}^{GR} - \hat{\theta}_{ytm}^{GR}) \qquad (4.1a)$$

$$= \hat{\theta}_{yt}^{GR} + K(\hat{\theta}_{y^{\prime}t^{\prime}}^{AK} - \hat{\theta}_{y^{\prime}t^{\prime}m^{\prime}}^{GR}) - (K-A)(\hat{\theta}_{yt}^{GR} - \hat{\theta}_{ytm}^{GR}) \qquad (4.1b)$$

The coefficients A and K are chosen to minimize the variance of $\hat{\theta}_{yt}^{AK}$. Thus, 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 $(\boldsymbol{\theta}_{xt} - \hat{\boldsymbol{\theta}}_{xt}^{HT})$, and in the second step, optimal regression is used to combine $\hat{\theta}_{yt}^{GR}$ with two additional predictor zero functions $(\hat{\theta}_{y^{\prime}t^{\prime}}^{AK} - \hat{\theta}_{y^{\prime}t^{\prime}m^{\prime}}^{GR})$ and $(\hat{\theta}_{yt}^{GR} - \hat{\theta}_{ytm}^{GR})$. The K-composite estimator, a predecessor of AK, is a special case of AK by setting $A=0$. Thus, in the K-composite estimator, 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 as 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.

Alternative estimators using MR were developed by Singh and Merkouris (1995). Since θ_{yt} is a linear parameter, MR for the linear case can be used here. The MR-composite is basically a GR-type estimator obtained by regressing (under a working covariance) the zero function $(\hat{\theta}_{yt}^{HT} - \theta_{yt})$ on all the predictor zero functions simultaneously, the usual predictors $(\boldsymbol{\theta}_{xt} - \hat{\boldsymbol{\theta}}_{xt}^{HT})$ and the additional ones $(\hat{\theta}_{y^{\prime}t^{\prime}}^{MR} - \hat{\theta}_{y^{\prime}t^{\prime}m^{\prime}}^{HT})$ and $(\hat{\theta}_{yt}^{HT} - \hat{\theta}_{ytm}^{HT})$. Notice that in this formulation, it is easy to make it multivariate, that is, additional predictors corresponding to a set of key study variables (y, z, etc.) can be used together to take advantage of the contemporaneous correlation between variables. This represents a departure from the traditional AK-estimator which is univariate. There is another aspect in which MR-composite departs from AK. Since the weight calibration is used for computing MR, all predictor zero functions should be in the form of auxiliary control total minus HT-estimator of the auxiliary variable at the current time t. Now, the additional predictor $(\hat{\theta}_{y^{\prime}t^{\prime}}^{MR} - \hat{\theta}_{y^{\prime}t^{\prime}m^{\prime}}^{HT})$ involving the variable $y^{\prime}$ from $t^{\prime}$ is not of this form. So this is transformed (compare with $\boldsymbol{\Psi}_*$ of (3.5)) to the zero function $(\tilde{\theta}_{y^{\prime}t}^{MR} - \hat{\theta}_{y^{\prime}tm}^{HT})$ where both $\tilde{\theta}_{y^{\prime}t}^{MR}$ and $\hat{\theta}_{y^{\prime}tm}^{HT}$ are (approximately) unbiased estimates of the conceptual parameter $\theta_{y^{\prime}t}$, i.e. the population total of $y^{\prime}$ at time t if the variable y' were assigned conceptually to all the population units at t. The estimate $\tilde{\theta}^{MR}_{y't}$ is obtained from $\hat{\theta}^{MR}_{y't}$ by calibrating the time t' MR-weights with respect to time t GR-auxiliary controls. On the other hand, the estimate $\hat{\theta}^{HT}_{y'tm}$ is obtained by augmenting the unit-level information about y' for the matched subsample via micromatching. For micromatching, some form of 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. Impact of this imputation, however, will be negligible if the probability that a time t-respondent is nonrespondent at t' and that its y-value has changed is small.

In computing MR-composite, the (random) control totals $\tilde{\theta}^{MR}_{y't}$ are treated as fixed, but their variability is taken into account when computing variance of $\hat{\theta}^{MR}_{yt}$. Also, to avoid instability due to too many predictors, Singh-Merkouris used only $(\tilde{\theta}^{MR}_{y't} - \hat{\theta}^{HT}_{y'tm})$ as additional predictors. They found good efficiency gains relative to the AK-estimator in the context of the Canadian Labour Force Survey. Other versions of MR-composite may be obtained by using the differences $[(\tilde{\theta}^{MR}_{y't}-\hat{\theta}^{HT}_{y'tm})-(\hat{\theta}^{HT}_{yt}-\hat{\theta}^{HT}_{ytm})]$ as additional predictors as in the K-estimator, or after this step, the remaining predictors $(\hat{\theta}^{HT}_{yt} - \hat{\theta}^{HT}_{ytm})$ can be added in a second step along with the usual GR-predictors. These and other measures of enhancement are currently being investigated in collaboration with B.P. Kennedy and S. Wu.

It may be instructive to consider the form of the estimator $\hat{\theta}^{MR}_{yt}$ in the case of simple random samples of size n from two successive occasions with the overlap sample being of size n^o. Also assume for simplicity that the population size N does not change over time. Now, the auxiliary variables are $\psi_{1k} = 1$, $\psi_{2k} = \alpha y'_k I_{\{k\in m'\}}$, $\psi_{3k} = (\alpha I_{\{k\in m\}} - I_{\{k\in s\}} y_k)$, where α is the inverse of the overlap proportion, i.e., $\alpha = n/n^o$, and $m = m'$ are the matched subsamples at t and t'. The three predictor zero functions are

$$g_{\psi 1} = \textstyle\sum_s (1) h_k - N \qquad (4.2a)$$
$$g_{\psi 2} = \alpha(\textstyle\sum_{m'} y'_k h'_k) - \sum_{s'} y'_k h'_k \qquad (4.2b)$$
$$g_{\psi 3} = \alpha(\textstyle\sum_m y_k h_k) - \sum_s y_k h_k \qquad (4.2c)$$

where $h_k = h'_k = N/n$. Here $g_{\psi 1}$ happens to be identically zero. From (3.4) we have $\hat{\theta}^{MR}_{yt}$ =

$$(N/n)\textstyle\sum_s y_k - (\hat{\sigma}_{yy'},\ 0)\begin{bmatrix} \alpha\hat{\sigma}^2_{y'} & (\alpha-1)\hat{\sigma}_{yy'} \\ (\alpha-1)\hat{\sigma}_{yy'} & (\alpha-1)\hat{\sigma}^2_{y} \end{bmatrix}^{-1}\begin{bmatrix} g_{\psi 2} \\ g_{\psi 3}\end{bmatrix} \qquad (4.3)$$

where $\hat{\sigma}_{yy'} = [\sum_m y_k y'_k - (\sum_m y_k)(\sum_{m'} y'_k)/n^o]/(n^o-1)$, and $\hat{\sigma}^2_{y'}$, $\hat{\sigma}^2_y$ are similarly defined. Note that the estimator $\hat{\theta}^{MR}_{yt}$ is not the optimal regression estimator because the estimator $\sum_{s'} y'_k h'_k$ in $g_{\psi 2}$ is treated as a fixed population control. The only difference between $\hat{\theta}^{MR}_{yt}$ and the optimal regression estimator is that the second diagonal element in the 2×2 matrix of (4.3) is changed to $\alpha\hat{\sigma}^2_y$ for the case of optimal regression. However, if $g_{\psi 3}$ is not used as a predictor, then $\hat{\theta}^{MR}_{yt}$ will coincide with the corresponding optimal regression estimator.

4.2 **Multi-frame Estimation by MR**

Here the problem of estimation involves combining information in independent samples from overlapping frames which together cover the target population. For simplicity, we consider the case of only two frames A and B, say. Let a, b, and c denote respectively the three nonoverlapping domains, $A \cap B^c$, $A^c \cap B$, and $A \cap B$. The pioneering work in this area is due to Hartley (1962, 1974). Hartley's estimator, $\hat{\theta}^H_y$, turns out to be a special case of the well known estimator of Fuller and Burmeister (1972) which is defined as $\hat{\theta}^{FB}$

$$=(\hat{\theta}^{GR}_{ya} + \hat{\theta}^{GR}_{yb} + \hat{\theta}^{GR}_{ycA}) - \beta_1(\hat{\theta}^{GR}_{ycA} - \hat{\theta}^{GR}_{ycB}) - \beta_2(\hat{N}^{GR}_{cA} - \hat{N}^{GR}_{cB}) \qquad (4.4)$$

where $\hat{\theta}^{GR}_{ycA}$, for example, denotes the GR-estimator of θ_{yc} based on sample from frame A, and β_1, β_2, are chosen to get optimal regression. If the predictor zero function $(\hat{N}^{GR}_{cA} - \hat{N}^{GR}_{cB})$ is dropped from (4.4), then $\hat{\theta}^{FB}_y$ reduces to $\hat{\theta}^H_y$. An alternative estimator was proposed recently by Skinner and Rao (1996) using the PML method.

The MR-estimator for multiframe problems was developed by Singh and Wu (1996). It is a GR-type solution with four types of predictor zero functions: the usual predictors $\hat{\boldsymbol{\theta}}^{HT}_{xA} - \boldsymbol{\theta}_{xA}$ from A, (ii) the usual predictors $\hat{\boldsymbol{\theta}}^{HT}_{xB} - \boldsymbol{\theta}_{xB}$ from B, (iii), the additional predictors $\hat{\theta}^{HT}_{ycA} - \hat{\theta}^{HT}_{ycB}$ from the samples in the common domain; these include $\hat{N}^{HT}_{cA} - \hat{N}^{HT}_{cB}$ for the counting variable as well as those for other selected study variables (thus rendering $\hat{\theta}^{MR}_{yt}$ multivariate in nature), and (iv) the additional predictors $\hat{\boldsymbol{\theta}}^{HT}_{xa} + \hat{\boldsymbol{\theta}}^{HT}_{xb} + (\hat{\boldsymbol{\theta}}^{HT}_{xcA} + \hat{\boldsymbol{\theta}}^{HT}_{xcB})/2 - \boldsymbol{\theta}_x$ for the combined frame. However, $\hat{\theta}^{MR}_{yt}$ differs from the single frame GR in some ways. The working covariance matrix is modified by replacing $\boldsymbol{W} = \textit{block diag}\,(\boldsymbol{W}_A, \boldsymbol{W}_B)$ by $\boldsymbol{\Lambda W}$ (compare with W_* of (3.5)) defined as

$$\boldsymbol{\Lambda W} = \textit{block diag}\,(\lambda_A \boldsymbol{W}_A, \lambda_B \boldsymbol{W}_B), \qquad (4.5)$$

where λ_A, λ_B are relative measures of the inverse effective sample size for the two sample designs from A and B. If the two designs are identical, then clearly $\lambda_A = \lambda_B = 1$. In practice, design effects can be used to estimate λ's.

The weight calibration is defined slightly differently from the usual GR as follows.

$$\hat{\theta}^{MR}_y = \hat{\theta}^{MR}_{ya} + \hat{\theta}^{MR}_{yb} + \hat{\theta}^{MR}_{yc} \qquad (4.6a)$$
$$= \boldsymbol{y}'_a \boldsymbol{w}_a + \boldsymbol{y}'_b \boldsymbol{w}_b + (\boldsymbol{y}'_{cA}\boldsymbol{w}_{cA} + \boldsymbol{y}'_{cB}\boldsymbol{w}_{cB})/2 \qquad (4.6b)$$

where

$$\boldsymbol{y}' = (\boldsymbol{y}'_a, \boldsymbol{y}'_{cA}, \boldsymbol{y}'_{cB}, \boldsymbol{y}_b)',\ \boldsymbol{w}' = (\boldsymbol{w}'_a, \boldsymbol{w}'_{cA}, \boldsymbol{w}'_{cB}, \boldsymbol{w}'_b) \qquad (4.6c)$$

In the special case of simple random samples from the two frames, $\hat{\theta}^{MR}_y$ coincides with the optimal regression estimator. For example, with the three predictors $\hat{N}^{HT}_A - N_A$, $\hat{N}^{HT}_B - N_B$, $\hat{N}^{HT}_{cA} - \hat{N}^{HT}_{cB}$, we can define the auxiliary variables as $\psi_{1k} = I_{\{k\in s(A)\}}$, $\psi_{2k} = I_{\{k\in s(B)\}}$, $\psi_{3k} = I_{\{k\in s(cA)\}} - I_{\{k\in s(cB)\}}$ where $s(cA)$, for example, is part of the sample $s(A)$ from A which is in c. The predictor zero functions can be written as

$$g_{\psi 1} = \textstyle\sum_{s(A)} (1) h_{kA} - N_A \qquad (4.7a)$$

$$g_{\psi 2} = \sum_{s(B)}(1)\, h_{kB} - N_B \qquad (4.7b)$$

$$g_{\psi 3} = \sum_{s(cA)}(1)\, h_{kA} - \sum_{s(cB)}(1)\, h_{kB}, \qquad (4.7c)$$

where $h_{kA}=N_A/n_A$, $h_{kB}=N_B/n_B$. Here $\lambda_A=n_A^{-1}$, $\lambda_B=n_B^{-1}$.

Now, from (4.6), $\hat{\theta}_y^{MR}$ can be easily computed as

$$\hat{\theta}_y^{MR} = (N_A/n_A)[\sum_{s(a)} y_k + \sum_{s(cA)} y_k/2] + \qquad (4.8a)$$
$$(N_B/n_B)[\sum_{s(b)} y_k + \sum_{s(cB)} y_k/2] - \hat{\beta}[(N_A/n_A)n_{cA} - (N_B/n_B)n_{cB}],$$

where $\hat{\beta} = [N_c(n_A^{-1*}\hat{N}_a + n_B^{-1*}\hat{N}_b)]^{-1} \times$
$[N_c(n_A^{-1*}\hat{\theta}_{ya} - n_B^{-1*}\hat{\theta}_{yb}) - (\hat{\theta}_{yc}/2)(n_A^{-1*}\hat{N}_a - n_B^{-1*}\hat{N}_b)]$,
and where

$$n_A^{-1*} = n_A^{-1} - N_A^{-1}, \quad n_B^{-1*} = n_B^{-1} - N_B^{-1}. \qquad (4.8b)$$

This is the optimal regression estimator.

4.3 Multi-phase Estimation by MR

We will consider the case of only two phases for simplicity. Two phase sampling is often conducted to collect inexpensive information about correlated auxiliary variables based on a large first phase sample, which is useful at the design as well as the estimation stage of the second phase sample. This problem of combining information is similar in principle to that of composite estimation. There exist GR-type estimators for this problem due to Särndal and Swensson (SS for short, 1987), Breidt and Fuller (BF for short, 1993), Armstrong and St-Jean (AS for short, 1994), and Hidiroglou and Särndal (HS for short, 1996), the latter two were motivated as calibration estimators. Alternative estimators can be developed using MR and are currently being investigated in collaboration with B. Tang.

Denoting by z -the auxiliary variables observed from the first phase sample $s(1)$, y -the study variables observed from the second phase sample $s(2)$, and by x -the usual auxiliary variables with known population totals $\boldsymbol{\theta}_x$, the predictor zero functions for combining information are $\boldsymbol{g}_{\psi(1)} = \hat{\boldsymbol{\theta}}_{x(1)}^{HT} - \boldsymbol{\theta}_x$, and $\boldsymbol{g}_{\psi(2)} = \hat{\boldsymbol{\theta}}_{x*(2)}^{HT} - \hat{\boldsymbol{\theta}}_{x*(1)}^{HT}$, where $x*$ includes both x- and z-variables, and the subscript (2), for example, denotes that the estimator is based on $s(2)$. Note that the HT-estimator based on phase 2 sample is not the usual one as it involves weights $h_k = h_{1k}h_{2k}$, h_{1k}^{-1} are the inclusion probabilities for phase 1 sample and h_{2k}^{-1} are the conditional inclusion probabilities for phase 2 given the phase 1 sample.

Assuming that y-values are known for the phase 1 sample, SS first consider a difference estimator using the predictor $\boldsymbol{g}_{\psi(1)}$, and then estimate this first phase sample parameter by another difference estimator using the predictor $\boldsymbol{g}_{\psi(2)}$. Thus, they propose a GR-type estimator by estimating the difference coefficients using superpopulation models. When dealing with several phases, BF propose reducing the set of predictors for the later phase by retaining only the most recent predictors from the earlier phase, the premise being that the recent predictors are better. This implies, in particular, a two step procedure for two phase sample, where first GR-estimators $\hat{\boldsymbol{\theta}}_{z(1)}^{GR}$ of $\boldsymbol{\theta}_z$ based on $s(1)$ using the auxiliary variable x are obtained, and then these are used as (random) population controls for $s(2)$ data with the initial weights $\{h_k\}$ to obtain a GR-type estimator for θ_y. Now using (3.5), a BF-type estimator can be defined using all the population controls, the matrix $\boldsymbol{W}$ in this case being $\mathrm{diag}(h_k, k \in s(2))$; the calibrated weights $\boldsymbol{w}_*$ take the form $h_k a_k^{BF}$ where a_k^{BF} are the adjustment factors. Note that, as in the case of composite estimation, the random population controls $\hat{\boldsymbol{\theta}}_{z(1)}^{GR}$ will be treated as fixed for defining the working covariance. However, for variance estimation, their variability should be accounted for. Like BF, AS propose a two step calibration except that at the second step they calibrate the weights $h_k a_{1k}$ to obtain $h_k a_{1k} a_{2k}^{AS}$ such that all the controls $\hat{\boldsymbol{\theta}}_{x*(1)}^{GR}$ are met; a_{1k}, a_{2k}^{AS} denote respectively the adjustment factors for the two steps. HS propose a somewhat different two step calibration in that at the second step a modified regression distance function is used to obtain $h_k(a_{1k}+a_{2k}^{HS}-1)$ from $h_k a_{1k}$ such that the same controls are met. Note that AS use $h_k a_{1k}$ in the denominator of the regression-distance while HS use h_k.

All the above estimators are GR-type. However, they do not give rise to optimal regression (OR) in general for simple random samples (SRS) at each phase. With MR, one can ensure (asymptotic) equivalence to OR for SRS by choosing the working covariance suitably. We can define the MR-estimator also as a 2-step calibration estimator: the first step is the same as in AS (i.e., obtain $h_{1k}a_{1k}$ from h_{1k} such that the controls $\boldsymbol{\theta}_x$ are met), and in the second step, a conditional regression calibration for $s(2)$ given $s(1)$ is performed to obtain $h_{2k}a_{2k}$ from h_{2k} such that the controls $\hat{\boldsymbol{\theta}}_{x*(1)}^{GR}$ are met. Note that in the second step of calibration, the usual h_k is replaced by h_{2k} (compare with $\boldsymbol{h}_*$ of (3.5)) . Also the predictors $\boldsymbol{g}_{\psi(2)}$ are transformed (compare with $\boldsymbol{\Psi}_*$ of (3.5)) to new (but asymptotically equivalent) predictors $\hat{\boldsymbol{\theta}}_{x*(2)}^{AHT} - \hat{\boldsymbol{\theta}}_{x*(1)}^{GR}$, where AHT denotes the adjusted HT-estimator when the adjusted weights $h_k a_{1k}$ are used. This allows for a simplified expression for the calibrated weights given by $h_k a_{1k} a_{2k}$.

In the special case of SRS when a_{1k} are simply constants (e.g., when the x-predictors consist of only the population counting variable), all the above estimators coincide with the OR estimator. To see this, let $\psi_{k1} = I_{k \in s(2)}$, $\psi_{k2} = z_k I_{k \in s(2)}$, we have the predictor zero functions:

$$g_{\psi 1(2)} = \sum_{s(2)}(1)\, h_k - N \qquad (4.9a)$$

$$g_{\psi 2(2)} = \sum_{s(2)} z_k h_k - \sum_{s(1)} z_k h_{1k} \qquad (4.9b)$$

where $h_k = N/n_2$, $h_{1k} = N/n_1$, n_1, n_2 are respectively the sample sizes for the two phases. From (3.5), we have, with the usual estimates of σ_{yz}, σ_z^2,

$$\hat{\theta}_y^{MR} = (N/n_2) \sum_{s(2)} y_k - (\hat{\sigma}_{yz}/\hat{\sigma}_z^2)\, g_{\psi 2(2)}. \qquad (4.10)$$

However, if a stratified SRS is used at phase 1, then a_{1k} will not be a constant, and the existing methods will not be equivalent to OR. This is also the case if phase 1 sample is used for stratification in phase 2. MR, on the other hand, remains asymptotically equivalent to

OR. It may be remarked that for multi-frame multiphase surveys, MR can also be defined by using MR-multiframe within each phase.

4.4 Median Estimation by MR

This is an example of nonlinear finite population parameter estimation. In the presence of auxiliary information $\boldsymbol{\theta}_x$, a commonly used method is to estimate median from the regression calibrated sampling weights. MR for nonlinear parameters (see section 3.1) also gives rise to this method. Note however that a different set of predictors may be desirable for nonlinear parameters, e.g., if the population median $\tilde{\theta}_x$ of x is also known, then $I\{x_k \le \tilde{\theta}_x\}$ would be preferable to the variable x_k used for linear parameters. As mentioned in Section 3.1, justification of MR is based on the method of estimating (or generalized zero) functions which requires the approximate representation (3.6). In the case of median, validity of this representation (with $f(\theta)$ as the density function evaluated at the median θ) follows from Francisco and Fuller's (1991) result on Bahadur representation for finite population quantiles. The sandwich formula (2.8) can be used to estimate the variance, see also Binder and Kovacevic (1995). As an alternative to the above simple estimator of median, more sophisticated estimators are among others due to Chambers and Dunstan (1986), and Rao, Kovar, and Mantel (1990). In light of the empirical study of Silva and Skinner (1995), the simple estimator, however, is expected to perform reasonably well. These and other related issues are currently being investigated in collaboration with M.S. Kovacevic.

4.5 Small Area Estimation by MR

This problem arises when the usual direct survey estimates for small areas which use information only from sample units within the small area are not stable due to small sample sizes. The problem is often addressed by using combined estimators which are linear combinations of synthetic and direct estimators. The synthetic component borrows strength from other areas or sources such as past data. In the linear combination, the direct estimator is shrunk toward the synthetic estimator by means of a predetermined shrinkage factor. The choice of this factor is based on consideration of the relative impact of bias in the synthetic component against the large variance of the direct component. Ghosh and Rao (1994) provide an excellent appraisal of various methods. In the design-based context, popular estimators are the sample size dependent estimator (SSD*) of Drew, Singh, and Choudhry (1982) and its modification (to be denoted by SSD) by Särndal and Hidiroglou (1989). The SSD* estimator for small area (or domain) d is given by

$$\hat{\theta}_{yd}^{SSD*} = \lambda_d \hat{\theta}_{yd}^{DIR} + (1-\lambda_d)\hat{\theta}_{yd}^{SYN} \qquad (4.11)$$

where $\hat{\theta}_{yd}^{DIR}$ is the GR-estimator $\hat{\theta}_{yd}^{HT} + \hat{\beta}_d'(\boldsymbol{\theta}_{xd} - \hat{\boldsymbol{\theta}}_{xd}^{HT})$, with $\hat{\beta}_d'$ as $\boldsymbol{y}_d'\boldsymbol{W}_d\boldsymbol{X}_d(\boldsymbol{X}_d'\boldsymbol{W}_d\boldsymbol{X}_d)$, $\hat{\theta}_{yd}^{SYN}$ is the regression-synthetic estimator given by $\hat{\boldsymbol{\beta}}_s'\boldsymbol{\theta}_{xd}$ with $\hat{\boldsymbol{\beta}}_s'$ as $\boldsymbol{y}_s'\boldsymbol{W}_s\boldsymbol{X}_s(\boldsymbol{X}_s'\boldsymbol{W}_s\boldsymbol{X}_s)^{-1}$, and the factor λ_d depends on the observed sample size (n_d) and the desired sample size (n_o) such that it is 1 if $n_d \ge n_0$, and n_d/n_0 otherwise. Note that the subscript d in $\hat{\boldsymbol{\beta}}_d$, for example, denotes that only the domain data are used, while the subscript s in $\hat{\boldsymbol{\beta}}_s$ denotes that the full sample data are used.

The SSD estimator modifies SSD* by using a modified direct estimator, $\hat{\theta}_{yd}^{MOD.DIR}$, in that $\hat{\boldsymbol{\beta}}_d$ is replaced by $\hat{\boldsymbol{\beta}}_s$. Thus

$$\hat{\theta}_{yd}^{SSD} = \lambda_d \hat{\theta}_{yd}^{MOD.DIR} + (1-\lambda_d)\hat{\theta}_{yd}^{SYN} \qquad (4.12a)$$

$$= \lambda_d \hat{\theta}_{yd}^{HT} + \hat{\boldsymbol{\beta}}_s'(\boldsymbol{\theta}_{xd} - \lambda_d \hat{\boldsymbol{\theta}}_{xd}^{HT}). \qquad (4.12b)$$

This is clearly related to the expression (3.9a) of $\hat{\theta}^{MR}$ for MR estimator in the presence of prior information θ^* about θ, θ^* being the synthetic estimator. (Note that the form of the SSD estimator can be justified by assuming a random small area effect.) This observation was used by Singh and Mian (1995) to propose generalized SSD (GSSD) estimators using the MR methodology. The GSSD estimator enhances SSD in several ways. For example, aggregate level constraints for small area estimators, in that the direct survey estimators ($\hat{\theta}_y^{DIR}$) for broad areas should coincide with the sum of small area estimators, can be easily incorporated via an additional zero functions $\hat{\theta}_y^{DIR} - \sum_{d'}\hat{\theta}_{yd'}^{HT}$ with appropriate working covariance with other zero functions. This condition can be viewed as part of the auxiliary information, and is clearly desirable in practice. By contrast, SSD estimators are further adjusted in a separate step to satisfy this condition, and consequently, the final SSD estimators have no longer the shrinkage form.

Another important enhancement by GSSD is based on the observation that SSD of (4.12b) can be computed by a modified version of the regression weight calibration method. To motivate this, one can visualize performing a mass imputation for each small area by borrowing data on the study and auxiliary variables (along with their sampling weights) from other areas so that the sample size n_d is boosted to n_d^* (say), and the total sample n is boosted to n^*. The sampling weights of the imputed data for each small area can be adjusted so that the total weight matches the area population. Now, for weight calibration, a penalized chi-square type distance function can be minimized under linear constraints to obtain a set of final weights for GSSD. The penalty component corresponds to the prior information.

There are important useful and practical consequences of being able to use weight calibration for computing GSSD. In particular, one could borrow data from sources other than the current sample such as past surveys for the same area. Moreover, for several key study variables, one could put aggregate level constraints simultaneously so that a single set of final calibrated weights satisfy all the required conditions and can be used for providing estimates for all study variables. This will obviously be very convenient in practice for a production system in meeting user demands, and for maintaining internal consistency of estimates.

4.6 Outlier-Prone Domain Estimation by MR

This problem is similar to that of small area estimation because of the instability of the direct estimator except that the instability is due to the presence of outlying or influential observations which in effect implies that the sample size is small. It is assumed that the outliers present in the sample can be identified and are representative. It is also assumed as in the case of small area estimation that conceptually there exists a random outlier-prone domain effect such that the prior estimate θ^* of θ can be taken as a regression-synthetic estimator. Now, MR method for small area estimation can be adapted for the outlier problem. The basic idea is to boost the sample by borrowing data on the study variable (y) and auxiliary variables ($\boldsymbol{x}$) along with their sampling weights from a suitable source (such as past surveys or administrative data) so that the observations detected as outliers before are no longer outliers. The sampling weights of borrowed (or mass imputed) data may be adjusted to satisfy the total population count. Note that the above method of borrowing strength to reduce the influence of outliers is in principle similar to the method of weight reduction commonly used for treating outliers. Now, the regression-synthetic estimator and the modified direct estimator can be obtained as in the case of SSD estimator. The outlier-prone domain estimator $\hat{\theta}_{yd}^{MR*}$ for domain d is defined as a combined estimator (4.12) where the shrinkage factor λ_d can now be obtained under a working rule as

$$\lambda_d = 1 - \sum_{s(d^o)} y_k h_k \,/ \sum_{s(d)} y_k h_k, \qquad (4.13)$$

where $s(d^o)$ is the subsample containing outliers. Note that λ_d depends on y and it tends to 1 as $n \to \infty$ because it is assumed that the outlier problem vanishes for large samples. This implies, in particular, that $\hat{\theta}_{yd}^{MR*}$ will be asymptotically design consistent. Now as in GSSD, a modified regression-weight calibration method can be used for computing the estimator $\hat{\theta}_{yd}^{MR*}$. However, in order to use final calibrated weights for other variables, the shrinkage factor should not depend on y. A way out is to define a conservative λ_d as min $\{\lambda_d(y)\}$ over a selected set of y-variables. Another possibility is to set $y_k = 1$ in (4.13). This will be reasonable if the outlier problem is due to large sampling weights and not extreme y-values. Moreover, when dealing with several domains, one can easily incorporate, as in the case of GSSD, aggregate level constraints so that domain estimates $\hat{\theta}_{yd}^{MR*}$ add up to the direct estimators $\hat{\theta}_y^{GR}$ at an aggregate level for a set of selected y-variables. The resulting final calibrated weights can then be used for producing domain estimates for all (selected or not) y-variables.

The problem of treatment of outliers has a long history. In the context of finite population estimation, there exist several alternative methods which are among others due to Hidiroglou and Srinath (1981) who used the idea of weight reduction, and Chambers (1986), Gwet and Rivest (1992) and Lee (1991) who used the idea of Huber's M-estimation, to propose outlier-robust estimators. For a good review of various methods, see Lee (1995). Alternative estimators based on the MR methodology and their comparison with other methods of outlier treatment are currently being investigated in collaboration with H. Lee.

5. CONCLUDING REMARKS

Using a design-based framework of zero functions in survey sampling, the MR method of estimation was proposed as a generalization of the familiar GR method. MR generalizes GR in several ways: (i) it provides a class of estimates which includes GR for a particular choice of predictors and the working covariance, (ii) it allows for general predictor zero functions which may be difference of two estimates, (iii) it encompasses estimation of nonlinear parameters such as median, (iv) it can incorporate prior information about parameters, and (v) as in GR, it continues to provide a set of final calibrated weights which can be used for all study variables. It was shown that MR may provide an important statistical technique for various problems related to combining information such as composite estimation for repeated surveys, estimation for multi-frame and multi-phase surveys, small area and outlier-prone domain estimation as well as in estimation of nonlinear parameters.

Acknowledgement The author is grateful to Dr. M.P. Singh and Prof. J.N.K. Rao for their interest, encouragement and helpful comments. This research was partially supported by a grant from Natural Sciences and Engineering Research Council of Canada held at Carleton University.

REFERENCES

Armstrong, J.B. and St-Jean, H. (1994). Generalized regression estimation for two-phase sample of tax records. *Survey Methodology*, **20**, 91-105.

Binder, D.A. (1983). On the variances of asymptotically normal estimators from complex surveys. *International Statistical Review*, **51**, 279-292.

Binder, D.A., and Kovacevic, M.S. (1995). Estimating some measures of income inequality from survey data: an application of the estimating equations approach. *Survey Methodology*, **21**, 137-145.

Breidt, F.J. and Fuller, W.A. (1993). Regression weighting for multiphase samples. *Sankhya, Ser. B*, **55**, 297-309.

Chambers, R.L. (1986). Outlier robust finite popula tion estimation. *Journal of the American Statistical Association*, **81**, 1063-1069.

Chambers, R.L. and Dunstan, R. (1986). Estimating distribution function from survey data. *Biomerika*, **73**, 597-604.

Deville, J.-C. and Särndal, C.-E. (1992). Calibration estimators in survey sampling. *Journal of the*

American Statistical Association, **87**, 376-382.

Drew, J.D., Singh, M.P., and Choudhry, G.H. (1982). Evaluation of small area estimation techniques for the Canadian Labour Force Survey. *Survey Methodology*, **8**, 17-47.

Francisco, C.A., and Fuller, W.A. (1991). Quantile estimation with a complex survey design. *Annals of Statistics*, **19**, 454-469.

Fuller, W.A. (1975), Regression analysis for sample surveys. *Sankhya Series C*, **37**, 117-132.

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

Fuller, W.A. and Burmeister, L.F. (1972). Estimators for samples selected from two overlapping frames *ASA Proc. Soc. Statist. Sec*, 245-249.

Ghosh, M. and Rao, J.N.K. (1994). Small area estimation: an appraisal (with discussion). *Statistical Science*, **9**, 55-93.

Godambe, V.P. (1966). A new approach to sampling from finite populations. *J. Roy. Statist. Soc. Ser. B*, **28**, 310-328.

Godambe, V.P., and Thompson, M.E. (1989). An extension of quasi-likehood estimation (with discussion). *Jour. Statistical Planning and Inference*, **22**, 137-172.

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

Gwet, J.-P., and Rivest, L.-P. (1992). Outlier resistant alternatives to the ratio estimator. *Jour. American Statistical Association*, **87**, 1174-1182.

Hartley, H.O. (1962). Multiple frame surveys. *ASA Proc. Soc. Statist. Sec.*, 203-206.

Hartley, H.O. (1974). Multiple frame methodology with selected applications. *Sankhya, C*, **36**, 99-118.

Henderson, C.R. (1975). Best linear unbiased estimation and prediction under a selection model. *Biometrics*, **31**, 423-447.

Hidiroglou, M.A., and Särndal, C.-E. (1996). Use of auxiliary information for two phase sampling. Technical report, Business Survey Methods Division, Statistics Canada, Ottawa.

Hidiroglou, M.A. and Srinath, K.P. (1981). Some estimators of a population total containing large units. *Jour. American Statistical Association*, **78**, 690-695.

Isaki, C.T., and Fuller, W.A. (1982). Survey design under the regression superpopulation model. *Jour American Statistical Association*, **77**, 89-96.

Lee, H. (1991). Model-based estimators that are robust to outliers. *Proc. Annual Res. Conf., Bureau of the Census*, 178-202.

Lee, H. (1995). Outliers in business surveys. In *Business Survey Methods* (ed. Cox, Binder, Chinnappa, Christianson, Colledge, Kott), 503-526.

Liang, K.-Y. and Zeger, S.L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, **73**, 13-22.

Nelder, J.A. and Wedderburn, R.W.H. (1972). Generalized linear models. *J. Roy. Statist. Soc. A*, **135**, 385-391.

Rao, C.R. (1968). A note on a previous lemma in the theory of least squares and some further results. *Sankya Ser. A*, **30**, 259-266.

Rao, J.N.K. (1994). Estimating totals and distribution functions using auxiliary information. *Jour. Off. Stat.*, **10**, 153-165.

Rao, J.N.K., Kovar, J.G. and Mantel, H. J. (1990). On estimating distribution functions and quantiles from survey data using auxiliary information. *Biometrika*, **77**, 365-375.

Rao, J.N.K. and Scott, A.J. (1981). The analysis of categorical data from complex sample surveys: chi-squared tests for goodness-of-fit and independence in two-way tables. *Jour. American Statistical Association*, **76**, 221-230.

Robinson, G.K, (1991). That BLUP is a good thing: the estimation of random effects (with discussion). *Statistical Science*, **6**, 15-51.

Särndal, C.-E. (1980). On π-inverse weighting versus best linear unbiased weighting in probability sampling. *Biometrika*, **67**, 639-650.

Särndal, C.-E. and Hidiroglou, M.A. (1989). Small domain estimation: a conditional analysis. *Jour. American Statistical Association*, **84**, 266-275.

Särndal, C.-E. and Swensson, B. (1987). A general view of estimation for two phases of selection with applications to two phase sampling and nonresponse. *Int. Statistical Review*, **55**, 279-294.

Silva, P.L.D.N. and Skinner, C.J. (1995). Estimating distribution functions with auxiliary information using poststratification. *Jour. Off. Statist.*, **11**, 271-294.

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

Singh, A.C. (1995). Predicting functions for generalization of BLUP to mixed nonlinear models. *ASA Proc. Biom. Sec.*

Singh, A.C. and Merkouris, P. (1995). Composite estimation by modified regression for repeated surveys. *ASA Proc. Surv. Res. Meth. Sec.*, 420-5.

Singh, A.C. and Mian, I.U.H. (1995). Generalized sample size dependent estimators for small areas. *Proc. Annual Res. Conf., U.S. Bureau of the Census*, 687-701.

Singh, A.C. and Wu, S. (1996). Estimation for multi frame complex surveys by modified regression. *SSC Proc. Surv. Meth. Sec.* (to appear)

Skinner, C.J. and Rao, J.N.K. (1996). Estimation in dual frame surveys with complex designs. *Jour. American Statistical Association*, **91**, 349-356.

Wedderburn, R.W.H. (1974). Quasi-likelihood functions, generalized linear models, and Gauss-Newton method. *Biometrika*, **61**, 439-447.