

A Model-Calibration Approach to Using Complete Auxiliary Information From Survey Data

Changbao Wu and Randy R. SITTER

Suppose that the finite population consists of N identifiable units. Associated with the i th unit are the study variable, y_i , and a vector of auxiliary variables, \mathbf{x}_i . The values $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N$ are known for the entire population (i.e., complete) but y_i is known only if the i th unit is selected in the sample. One of the fundamental questions is how to effectively use the complete auxiliary information at the estimation stage. In this article, a unified model-assisted framework has been attempted using a proposed model-calibration technique. The proposed model-calibration estimators can handle any linear or nonlinear working models and reduce to the conventional calibration estimators of Deville and Särndal and/or the generalized regression estimators in the linear model case. The pseudoempirical maximum likelihood estimator of Chen and Sitter, when used in this setting, gives an estimator that is asymptotically equivalent to the model-calibration estimator but with positive weights. Some existing estimators using auxiliary information are reexamined under this framework. The estimation of the finite population distribution function, using complete auxiliary information, is also considered, and estimators based on a general model are presented. Results of a limited simulation study on the performance of the proposed estimators are reported.

KEY WORDS: Distribution function; Empirical likelihood; Estimating equations; Generalized linear models; Generalized regression estimator; Superpopulation.

1. INTRODUCTION

In sample surveys, auxiliary information on the finite population is often used to increase the precision of estimators of the population mean, total or distribution function. In the simplest settings, ratio and regression estimators incorporate known finite population means of auxiliary variables. For more general situations, there have been three main methods proposed in the literature which can be categorized as model-assisted approaches: generalized regression estimators (GREG) (Cassel, Särndal, and Wretman 1976; Särndal 1980); calibration estimators (Deville and Särndal 1992); and more recently empirical likelihood methods (Chen and Qin 1993; Chen and Sitter 1999; Zhong and Rao 1996). All of these methods have only been discussed in the context of a linear regression working model and essentially incorporate the auxiliary variables through their known population means even when the auxiliary variables are known for every unit in the population.

Note that by model-assisted we mean that the estimators are approximately (asymptotically) design unbiased irrespective of whether the working model is correct or not, and are particularly efficient if a working model is correct. By design-unbiased we mean the estimator is unbiased over repeated sampling of units. We use the terms model and working model interchangeably keeping in mind that inferences and the asymptotic framework are design-based with the working-model only used to increase efficiency.

In this paper, we consider the use of more complex models in obtaining model-assisted estimators by generalizing the calibration method above. We term the approach *model calibration* for reasons that will become readily apparent. We argue that, under a general modeling process, complete auxiliary information should be incorporated into the construction of

estimators through fitted values. How to do this properly is fairly straightforward in the case of a GREG (see Section 3) but not so for calibration. We introduce a general framework from which to do this that is simple, and reduces to the usual estimators when a linear model is used.

Once this generalization is realized, some interesting relationships between a linear model and the use of complete auxiliary information become more obvious and are discussed. Also, some differences between the approaches become more distinct. For example, it has been noted that the calibration estimator reduces to a GREG under a chi-squared distance measure (Deville and Särndal 1992), where an underlying linear regression model is used. This is no longer the case when the methods are generalized to nonlinear models, and the proposed model-calibration method performs better.

In Section 2, we briefly review the calibration method and discuss its implicit model-assisted nature and relationship to a linear model. In Section 3, we propose a model-calibration method for incorporating auxiliary information into estimation of the population mean under a very general working model that includes linear and nonlinear regression and generalized linear models as special cases. We go on to show that the resulting estimator is asymptotically design-unbiased irrespective of the working model and reduces to the usual calibration estimator under a linear working model. Also in Section 3, we discuss the extension of the GREG and the pseudoempirical maximum likelihood methods to the general model and demonstrate that unlike in the linear model case, the extended calibration and the extended GREG do not yield the same estimator, even asymptotically. We go on to demonstrate, through a small simulation study, that the model-calibration estimator and the pseudoempirical likelihood method are superior. In Section 4, we discuss estimation of the finite population distribution functions and show that our approaches for the mean case also provide a unified framework for the estimation of distribution functions. Some concluding remarks are given in Section 5.

Changbao Wu is Assistant Professor, Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, ON, Canada, N2L 3G1. (Email: cbwu@uwaterloo.ca). Randy R. Sitter is Professor, Department of Statistics and Actuarial Science, Simon Fraser University, Burnaby, BC, Canada V5A 1S6. (E-mail: sitter@stat.sfu.ca). This research was supported by a grant from the Natural Sciences and Engineering Research Council of Canada. The authors acknowledge helpful comments from Dr. Jiahua Chen and constructive suggestions and comments from referees.

2. HOW THE USUAL CALIBRATION METHOD RELATES TO A LINEAR MODEL

Consider a finite population consisting of N identifiable units. Associated with the i th unit are, the study variable, y_i , and a vector of auxiliary variables, \mathbf{x}_i . The values $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N$ are known for the entire population (i.e., complete) but y_i is known only if the i th unit is selected in the sample, s . Assume the inclusion probabilities $\pi_i = \Pr(i \in s)$ are strictly positive. For the moment we restrict attention to estimating the population total $Y = \sum_{i=1}^N y_i$.

Deville and Särndal (1992) introduce the notion of a calibration estimator of Y , which is constructed as $\hat{Y}_C = \sum_{i \in s} w_i y_i$, where the calibration weights w_i 's are chosen to minimize their average distance Φ_s from the basic design weights, $d_i = 1/\pi_i$, that are used in the Horvitz–Thompson estimator $\hat{Y}_{HT} = \sum_{i \in s} d_i y_i$, subject to the constraint

$$\sum_{i \in s} w_i \mathbf{x}_i = \mathbf{X}, \tag{1}$$

where \mathbf{X} are the known population totals for the auxiliary variables. The distance measure Φ_s is most commonly chosen as

$$\Phi_s = \sum_{i \in s} (w_i - d_i)^2 / (d_i q_i), \tag{2}$$

where the q_i 's are known positive weights unrelated to d_i . The resulting calibration estimator is

$$\hat{Y}_C = \sum_{i \in s} w_i y_i = \hat{Y}_{HT} + (\mathbf{X} - \hat{\mathbf{X}}_{HT})' \hat{\mathbf{B}}, \tag{3}$$

where $\hat{\mathbf{X}}_{HT} = \sum_{i \in s} d_i \mathbf{x}_i$ and $\hat{\mathbf{B}} = \{\sum_{i \in s} d_i q_i \mathbf{x}_i \mathbf{x}_i'\}^{-1} \sum_{i \in s} d_i q_i \mathbf{x}_i y_i$. The uniform weights $q_i = 1$ are used in most applications, but unequal weights can also be motivated as in Example 1 of Deville and Särndal (1992). The calibrated weights, w_i , give perfect estimates when applied to the auxiliary variables. Deville and Särndal (1992) argue that “. . . weights that perform well for the auxiliary variable also should perform well for the study variable.” However, it is an implicit underlying assumption that y and \mathbf{x} are linearly related that makes this a valid argument. For example, in the case of scalar x and $\mathbf{x}' = (1, x_i)$ is used in (1), it is clear that $y_i = \beta_0 + \beta_1 x_i$ implies $\hat{Y}_C = Y$. If a curved relationship exists between y and x , the so constructed calibration estimator could be very inefficient. For instance, if $\log(y_i) = \beta_0 + \beta_1 x_i$, there is no compelling reason to use \hat{Y}_C .

The point we want to illustrate is that, it is the relationship between y and \mathbf{x} , hopefully captured by the working-model, that determines how the auxiliary information should best be used. In fact, Deville and Särndal (1992) show that, for any Φ_s , \hat{Y}_C is asymptotically equivalent to (3), which is the generalized regression estimator, \hat{Y}_{GR} , and the GREG is motivated as a model-assisted estimator using a linear working-model (Särndal 1980). Thus, \hat{Y}_C is implicitly relying on a linear relationship between y and \mathbf{x} without explicitly stating so. Another point relates to the issue of complete information on the \mathbf{x} variables (i.e., known for all units in the population) versus only knowing the value of their population totals, \mathbf{X} . The GREG is motivated by using the predicted values from a linear model for each \mathbf{x}_i , $i = 1, \dots, N$. However, the resulting estimator in (3) only needs \mathbf{X} to be implemented. As we will see, this is related to the use of a linear model.

3. MODEL-CALIBRATION ESTIMATOR OF THE MEAN

In this section, we propose a unified framework for the estimation of the finite population mean $\bar{Y} = Y/N$. We will use a model-assisted approach. That is, our estimators of \bar{Y} will be design-consistent regardless of the working-model used to construct it, but will be particularly efficient if the working model is correct. This can be accomplished by first using the (y_i, \mathbf{x}_i) for $i \in s$ to build the model and then calibrating to the predicted values from the model using:

(1) a direct calibration argument such as was discussed in the previous section;

(2) using a pseudoempirical likelihood approach (Chen and Sitter 1999); or

(3) using a generalized difference estimator (Cassel, Särndal, and Wretman 1976; Särndal 1980). We now briefly discuss the modeling step first and then consider these three methods of calibrating on the predicted values.

3.1 Modeling

Assume the relationship between y and \mathbf{x} can be described by a superpopulation model through the first and second moments,

$$E_\xi(y_i | \mathbf{x}_i) = \mu(\mathbf{x}_i, \boldsymbol{\theta}), \quad V_\xi(y_i | \mathbf{x}_i) = v_i^2 \sigma^2, \tag{4}$$

$i = 1, 2, \dots, N,$

where $\boldsymbol{\theta} = (\theta_0, \dots, \theta_p)'$ and σ^2 are unknown superpopulation parameters, $\mu(\mathbf{x}, \boldsymbol{\theta})$ is a known function of \mathbf{x} and $\boldsymbol{\theta}$, the v_i is a known function of \mathbf{x}_i or $\mu_i = \mu(\mathbf{x}_i, \boldsymbol{\theta})$, and E_ξ and V_ξ denote the expectation and variance with respect to the superpopulation model. We also assume that $(y_1, \mathbf{x}_1), \dots, (y_N, \mathbf{x}_N)$ are mutually independent.

The model structure (4) is quite general and includes two very important cases: (1) the linear or nonlinear regression model,

$$y_i = \mu(\mathbf{x}_i, \boldsymbol{\theta}) + v_i \varepsilon_i, \quad i = 1, 2, \dots, N, \tag{5}$$

where ε_i 's are independently and identically distributed random variables with $E_\xi(\varepsilon_i) = 0$ and $V_\xi(\varepsilon_i) = \sigma^2$, and $v_i = v(\mathbf{x}_i)$ is a strictly positive known function of \mathbf{x}_i only; and (2) the generalized linear model,

$$g(\mu_i) = \mathbf{x}_i' \boldsymbol{\theta}, \quad V_\xi(y_i | \mathbf{x}_i) = v(\mu_i), \tag{6}$$

$i = 1, 2, \dots, N,$

where $\mu_i = E_\xi(y_i | \mathbf{x}_i)$, $g(\cdot)$ is a link function and $v(\cdot)$ is a variance function.

Consider a design-based method for estimating the model parameters. When a model-based approach is employed, (y_i, \mathbf{x}_i) , $i \in s$ is viewed as an iid sample from the superpopulation. The superpopulation parameters, $\boldsymbol{\theta}$, can then be estimated using standard procedures. Under the design-based framework, the sample data may not follow the same model structure as that of the whole finite population under a complex sampling scheme and $\boldsymbol{\theta}$ may be meaningless from the design-based point of view. In this case, $\boldsymbol{\theta}$ is replaced by $\boldsymbol{\theta}_N$, an estimate

of θ based on the data from the entire population. θ_N is then estimated by $\hat{\theta}$, a design-based estimate from the sampled data (Godambe and Thompson 1986).

For illustration, consider two important cases.

Case I. θ_N can be expressed explicitly as functions of population totals for properly defined population variables. For example, under a linear regression model, θ_N is the regression parameter of the finite population: $\theta_N = (\mathbf{X}'_N \mathbf{X}_N)^{-1} \mathbf{X}'_N \mathbf{y}_N$, where \mathbf{X}_N is the $N \times (p+1)$ matrix with rows $(1, \mathbf{x}'_i)$ for $i = 1, \dots, N$ and $\mathbf{y}_N = (y_1, \dots, y_N)'$. A design-based estimator $\hat{\theta}$ is obtained by plugging in design-based estimates for various population totals in θ_N : $\hat{\theta} = (\mathbf{X}'_n \Pi^{-1} \mathbf{X}_n)^{-1} \mathbf{X}'_n \Pi^{-1} \mathbf{y}_n$, where $\Pi = \text{diag}(\pi_1, \dots, \pi_n)$ and \mathbf{X}_n and \mathbf{y}_n in obvious notation.

Case II. θ_N is defined by estimating equations. Suppose that the generalized linear model (6) is assumed. We define θ_N as the maximum quasilielihood estimator of θ based on the entire finite population, that is, the solution of the estimating equations (Molina and Skinner 1992):

$$\sum_{i=1}^N \mathbf{X}'_i g^{(-1)}\{\mu(\mathbf{x}_i, \theta)\} v^{-1}\{\mu(\mathbf{x}_i, \theta)\} [y_i - \mu(\mathbf{x}_i, \theta)] = 0, \quad (7)$$

where $\mathbf{X}'_i = (1, \mathbf{x}'_i)$ and $g^{(-1)}(u) = [dg(u)/du]^{-1}$. The estimating function on the left-hand side of (7) is a population total, $\hat{\theta}$ is defined as the solution of the design-based sample version of (7), that is, the solution of the estimating equations

$$\sum_{i \in s} d_i \mathbf{X}'_i g^{(-1)}\{\mu(\mathbf{x}_i, \theta)\} v^{-1}\{\mu(\mathbf{x}_i, \theta)\} [y_i - \mu(\mathbf{x}_i, \theta)] = 0.$$

The estimate $\hat{\theta}$ is then obtained by standard Newton–Raphson iterative procedures. Under certain regularity conditions (similar to those used by Binder 1983), it can be shown that in both Cases I and II, $\hat{\theta} = \theta_N + O_p(n^{-1/2})$ (Wu 1999).

3.2 Model Calibration

Under Model (4), auxiliary information should be used through the fitted values $\mu(\mathbf{x}_i, \theta)$, $i = 1, \dots, N$. To do this, we define the model-calibration estimator of \bar{Y} as $\hat{Y}_{MC} = N^{-1} \sum_{i \in s} w_i y_i$, where the calibrated weights, w_i , minimize an average distance between w_i 's and d_i 's, subject to

$$N^{-1} \sum_{i \in s} w_i = 1 \quad \sum_{i \in s} w_i \mu(\mathbf{x}_i, \hat{\theta}) = \sum_{i=1}^N \mu(\mathbf{x}_i, \hat{\theta}). \quad (8)$$

One should note that in the original formulation of the calibration estimator presented in the previous section, the constraint $N^{-1} \sum_{i \in s} w_i = 1$ is not present. If this constraint is added, the resulting estimator under no auxiliary information is $\hat{Y} = \sum_{i \in s} d_i y_i / \sum_{i \in s} d_i$ and not $\hat{Y}_{HT} = N^{-1} \sum_{i \in s} d_i y_i$. It was illustrated in Rao (1966) and later in the more well known Basu (1971) elephant example that even though the first estimator estimates the population size N and the second uses its known quantity, the first has better properties. This is true for calibration generally. This constraint arises quite naturally in the case of pseudoempirical maximum likelihood estimators (PEMLE) (Chen and Sitter 1999).

We restrict our discussion to the chi-squared distance given in (2). The resulting model-calibration estimator then follows

directly by minimizing (2) subject to (8) using a Lagrange multiplier approach as in Deville and Särndal (1992), and is given by

$$\hat{Y}_{MC} = \hat{Y}_{HT} + \left\{ N^{-1} \sum_{i=1}^N \hat{\mu}_i - N^{-1} \sum_{i \in s} d_i \hat{\mu}_i \right\} \hat{B}_N, \quad (9)$$

where $\hat{B}_N = \sum_{i \in s} d_i q_i (\hat{\mu}_i - \bar{\mu})(y_i - \bar{y}) / \sum_{i \in s} d_i q_i (\hat{\mu}_i - \bar{\mu})^2$, $\bar{y} = \sum_{i \in s} d_i q_i y_i / \sum_{i \in s} d_i q_i$ and $\bar{\mu} = \sum_{i \in s} d_i q_i \hat{\mu}_i / \sum_{i \in s} d_i q_i$. Note that under a design-based framework, B_N does not go to one.

If constraint $N^{-1} \sum_{i \in s} w_i = 1$ is dropped, the single calibration equation $\sum_{i=1}^N \mu(\mathbf{x}_i, \theta) = \sum_{i \in s} w_i \mu(\mathbf{x}_i, \theta)$ yields

$$\hat{Y}_{MC}^* = \hat{Y}_{HT} + \left\{ N^{-1} \sum_{i=1}^N \hat{\mu}_i - N^{-1} \sum_{i \in s} d_i \hat{\mu}_i \right\} \hat{B}_N^*, \quad (10)$$

where $\hat{B}_N^* = \sum_{i \in s} d_i q_i \hat{\mu}_i y_i / \sum_{i \in s} d_i q_i \hat{\mu}_i^2$.

The important properties of (9) and (10) are summarized in the following theorem. We assume that there is a sequence of sampling designs and a sequence of finite populations, indexed by ν . Both the sample size n_ν and the population size N_ν approach infinity as $\nu \rightarrow \infty$. All limiting processes are understood to be as $\nu \rightarrow \infty$, but the ν is suppressed to simplify notation.

The following conditions are necessary for Theorem 1. Some of the notations used are described in the Appendix.

- (i) $\hat{\theta} = \theta_N + O_p(n^{-1/2})$ and $\theta_N \rightarrow \theta$;
- (ii) for each \mathbf{x}_i , $\partial \mu(\mathbf{x}_i, \mathbf{t}) / \partial \mathbf{t}$ is continuous in \mathbf{t} and $|\partial \mu(\mathbf{x}_i, \mathbf{t}) / \partial \mathbf{t}| \leq h(\mathbf{x}_i, \theta)$ for \mathbf{t} in a neighborhood of θ , and $N^{-1} \sum_{i=1}^N h(\mathbf{x}_i, \theta) = O(1)$;
- (iii) the basic design weights, $d_i = 1/\pi_i$, satisfy that the Horvitz–Thompson estimators for certain population means are asymptotically normally distributed;
- (iv) for each \mathbf{x}_i , $\partial^2 \mu(\mathbf{x}_i, \mathbf{t}) / \partial \mathbf{t} \partial \mathbf{t}'$ is continuous in \mathbf{t} and $\max_{j,k} |\partial^2 \mu(\mathbf{x}_i, \mathbf{t}) / \partial t_j \partial t_k| \leq k(\mathbf{x}_i, \theta)$ for \mathbf{t} in a neighborhood of θ , and $N^{-1} \sum_{i=1}^N k(\mathbf{x}_i, \theta) = O(1)$.

Theorem 1. (1) Assume the working-model used to construct the estimators has general structure (4). Under conditions (i) ~ (iii) given above, both \hat{Y}_{MC} and \hat{Y}_{MC}^* equal $\hat{Y}_{HT} + O_p(n^{-1/2})$ and are thus asymptotically design-unbiased estimators for Y , irrespective of whether the model is correct or not. They are also both approximately model-unbiased under condition (i);

(2) Under conditions (i) ~ (iv) given previously, the asymptotic design-variance of \hat{Y}_{MC} is given by

$$V(\hat{Y}_{MC}) \doteq N^{-2} \sum_{i < j} (\pi_i \pi_j - \pi_{ij}) \left(\frac{U_i}{\pi_i} - \frac{U_j}{\pi_j} \right)^2, \quad (11)$$

where π_{ij} are second order inclusion probabilities, $U_i = y_i - \mu_i B_N$, $\mu_i = \mu(\mathbf{x}_i, \theta_N)$, $B_N = \sum_{i=1}^N q_i (\mu_i - \bar{\mu}_N)(y_i - \bar{Y}) / \sum_{i=1}^N q_i (\mu_i - \bar{\mu}_N)^2$, and $\bar{\mu}_N = N^{-1} \sum_{i=1}^N \mu_i$. $V(\hat{Y}_{MC})$ can be estimated by

$$v(\hat{Y}_{MC}) = N^{-2} \sum_{i < j} \frac{\pi_i \pi_j - \pi_{ij}}{\pi_{ij}} \left(\frac{u_i}{\pi_i} - \frac{u_j}{\pi_j} \right)^2, \quad (12)$$

where $u_i = y_i - \hat{\mu}_i \hat{B}_N$. Similar results hold for \hat{Y}_{MC}^* if we replace B_N by $B_N^* = \sum_{i=1}^N q_i \mu_i y_i / \sum_{i=1}^N q_i \mu_i^2$ and \hat{B}_N by \hat{B}_N^* ;

(3) If $q_i = 1/v_i^2$ in Φ_s , then both Y_{MC} and Y_{MC}^* reduce to the conventional calibration estimator (or GREG) when a linear working-model is used, where

$$\mu(\mathbf{x}_i, \boldsymbol{\theta}) = \theta_0 + \theta_1 x_{i1} + \dots + \theta_p x_{ip}. \tag{13}$$

Proof. See Appendix.

Thus, both \hat{Y}_{MC} and \hat{Y}_{MC}^* are model-assisted in this sense and can handle any linear or non-linear models. That is, they are both design-consistent irrespective of whether the model holds and particularly efficient if the model is correct. Also, in the case of no modeling error, (i.e., $y_i = \mu_i$), we have $\hat{Y}_{MC} = \hat{Y}_{MC}^* = \bar{Y}$. In addition, both Y_{MC} and Y_{MC}^* reduce to the conventional calibration estimator (Deville and Särndal 1992) (same as the GREG) under the linear working model given in (13). Note that this remains true for \hat{Y}_{MC}^* even if θ_0 is removed from (13) but not for \hat{Y}_{MC} .

3.3 Pemle

Chen and Sitter (1999) propose using a pseudoempirical maximum likelihood estimator (PEMLE), $\hat{Y} = \sum_{i \in s} \hat{p}_i y_i$ where the \hat{p}_i are obtained by maximizing

$$\hat{l}(\mathbf{p}) = \sum_{i \in s} d_i \log p_i, \tag{14}$$

which is a design-unbiased estimator of the log-empirical likelihood one would use if one had the entire population: $E_p(\sum_{i \in s} d_i \log p_i) = \sum_{i=1}^N \log p_i$. Here, E_p refers to expectation under the sampling design. For auxiliary information of the general form $N^{-1} \sum_{i=1}^N \mathbf{u}_i = 0$, with $\mathbf{u}_i = \mathbf{u}(y_i, \mathbf{x}_i)$ (possibly vector valued), the method then reduces to maximizing (14) subject to

$$\sum_{i \in s} p_i = 1, \quad \sum_{i \in s} p_i \mathbf{u}_i = 0 \quad (0 \leq p_i \leq 1). \tag{15}$$

Chen and Sitter (1999) then go on to primarily focus on estimating \bar{Y} with $\bar{\mathbf{X}}$ known, i.e. $\mathbf{u}(y_i, \mathbf{x}_i) = (\mathbf{x}_i - \bar{\mathbf{X}})$. The resulting estimator is asymptotically equivalent to a GREG discussed next (Note that for vector valued $\mathbf{x}_i - \bar{\mathbf{X}}$ the PEMLE proposed by Chen and Sitter involves a constrained high-dimensional maximization problem.) Thus much like the calibration method, there is implicit use of a linear relationship between y and \mathbf{x} .

To extend to model (4) we merely define scalar $u_i = u(y_i, \mathbf{x}_i) = \mu(\mathbf{x}_i, \boldsymbol{\theta}) - N^{-1} \sum_{i=1}^N \mu(\mathbf{x}_i, \boldsymbol{\theta})$. In applications $\boldsymbol{\theta}$ will be replaced by $\hat{\boldsymbol{\theta}}$. The PEMLE of \bar{Y} is then defined as $\hat{Y}_{EL} = \sum_{i \in s} \hat{p}_i y_i$, where \hat{p}_i 's maximize $\hat{l}(\mathbf{p})$ subject to (15).

Paralleling Chen and Sitter (1999), the Lagrange multiplier method can be used to show that $\hat{p}_i = w_i / (1 + \lambda u_i)$ for $i \in s$, and the scalar Lagrange multiplier, λ , is the solution to

$$\sum_{i \in s} \frac{w_i u_i}{1 + \lambda u_i} = 0, \tag{16}$$

where $w_i = d_i / \sum_{i \in s} d_i$. Note that no auxiliary information translates into $u_i = 0$ and the resulting PEMLE of \bar{Y} is $\hat{Y} =$

$\sum_{i \in s} d_i y_i / \sum_{i \in s} d_i$. One advantage to this approach is that the resulting weights are always positive, which may not be true for the other two methods. The intrinsic positive weights associated with this approach make the technique generally applicable to the estimation of distribution functions and quantiles.

A theorem analogous to theorem 1 of Chen and Sitter (1999) can then be proved.

Theorem 2. Under the same asymptotic framework and conditions (i) ~ (iii) assumed in Theorem 1, and (v) ~ (vii) given next,

$$\hat{Y}_{EL} = \hat{Y}_{MC} + o_p(n^{-1/2}).$$

Let $u_i = \mu(\mathbf{x}_i, \boldsymbol{\theta}_N) - N^{-1} \sum_{i=1}^N \mu(\mathbf{x}_i, \boldsymbol{\theta}_N)$, $h_i = h(\mathbf{x}_i, \boldsymbol{\theta}_N)$, where $h(\mathbf{x}_i, \boldsymbol{\theta}_N)$ is defined in condition (ii) of Theorem 1. The conditions needed are: (v) $u^* = \max_{i \in s} |u_i| = o_p(n^{1/2})$; (vi) $\sum_{i \in s} d_i u_i / \sum_{i \in s} d_i u_i^2 = O_p(n^{-1/2})$; (vii) $h^* = \max_{i \in s} |h_i| = o_p(n)$.

Proof. The proof of Theorem 2 follows directly from the proof of Theorem 1 in Chen and Sitter (1999) by noting that \hat{Y}_{MC} is asymptotically equivalent to a regression estimator treating $\mu(\mathbf{x}_i, \boldsymbol{\theta}_N)$ as a scalar auxiliary variable (see Section 3.2), with one extra step: show that the conditions 2) and 3) used by Chen and Sitter (1999) still hold when $\boldsymbol{\theta}_N$ in u_i is replaced by $\hat{\boldsymbol{\theta}}$. This can be done by applying a Taylor series expansion to $\mu(\mathbf{x}_i, \hat{\boldsymbol{\theta}})$ at $\hat{\boldsymbol{\theta}} = \boldsymbol{\theta}_N$ and using the extra conditions imposed here. Details of the proof can be found in Wu (1999).

Since \hat{Y}_{EL} is asymptotically equivalent to \hat{Y}_{MC} , the same analytical design variance formula and variance estimator of \hat{Y}_{MC} from Theorem 1 can be used for \hat{Y}_{EL} . However, it is more attractive to use resampling variance estimators in this case since the calculation is directly applied to \hat{Y}_{EL} . Chen and Sitter (1999) show that a jackknife variance estimator for PEMLE has better finite sample performance than the analytical one. In general, jackknife variance estimators also perform better conditionally, conditioning on the means of the \mathbf{x} variables of the sample configuration.

Discussion of conditions on u_i 's and how smoothness of the models might translate moment conditions on the x_i 's to these conditions with references to examples of different designs are discussed in the Appendixes of Chen and Sitter (1999).

3.4 Generalized Difference Estimator

The well-known generalized regression estimator (GREG) (Cassel, Särndal, and Wretman 1976; Särndal 1980) can be motivated as a model-assisted generalized difference estimator (GD) that uses a linear working-model. Suppose we assume a linear model as in (5) with $\mu_i = \mu(\mathbf{x}_i, \boldsymbol{\theta}) = \mathbf{x}_i' \boldsymbol{\theta}$. The GREG can then be written (Särndal 1980) as

$$\hat{Y}_{GD} = N^{-1} \left\{ \sum_{i \in s} d_i y_i - \sum_{i \in s} d_i \hat{\mu}_i + \sum_{i=1}^N \hat{\mu}_i \right\}, \tag{17}$$

where $\hat{\mu}_i = \mathbf{x}_i' \hat{\boldsymbol{\theta}}$. This estimator is obtained by choosing $a_i = \hat{\mu}_i$ in the usual design-based difference estimator $\hat{Y}_D =$

$N^{-1}\{\sum_{i \in S} \hat{d}_i y_i - \sum_{i \in S} d_i a_i + \sum_{i=1}^N a_i\}$. The choice of $a_i = \mu_i = E_{\xi}(y_i)$ in \hat{Y}_D is optimal in that it minimizes $E_p\{V_{\xi}(\hat{Y}_D - \bar{Y})\}$ if $\pi_i \propto v_i = [V_{\xi}(y_i)]^{1/2}$ (Cassel, Särndal, and Wretman 1976). Note that for $\hat{\mu}_i = \mathbf{x}_i' \hat{\boldsymbol{\theta}}$ (17) will depend on the \mathbf{x} -values only through $\hat{\mathbf{X}}_{HT}$ and $\hat{\mathbf{X}}$. It is now clear that this motivation generalizes naturally to more complex models by allowing $\hat{\mu}_i = \mu(\mathbf{x}_i, \hat{\boldsymbol{\theta}})$ to come from the development in Section 3.1. However, in this case the resulting estimator depends on the \mathbf{x} values in a more complicated way and in particular complete auxiliary information is necessary to apply the method. Though this extension is quite natural, we are unaware of reference to it in existing literature.

This generalization of \hat{Y}_{GR} to \hat{Y}_{GD} using model (4) shares many of the nice properties of \hat{Y}_{MC} . In particular: (i) Theorem 1 can be restated for \hat{Y}_{GD} with similar proof (replace \hat{B}_N and \hat{B}_N by 1 in the variance formula); (ii) if $y_i = \mu_i$, $\hat{Y}_{GD} = \bar{Y}$; and (iii) using a linear working model it reduces to the usual GREG.

It is interesting to note that \hat{Y}_{MC} and \hat{Y}_{GD} are not the same when the general model is used to construct them, as they are when a linear model is used. In fact, they are not even asymptotically the same under the design-based framework, because \hat{B}_N does not go to 1. The behavior of \hat{Y}_{GD} is associated with the ‘‘goodness’’ of approximation $y_i \doteq \mu(\mathbf{x}_i, \hat{\boldsymbol{\theta}})$. That is, the performance of \hat{Y}_{GD} depends largely on modeling variation. The reduction of the design-variance from using \hat{Y}_{GD} over \hat{Y}_{HT} depends on how variate the fitted residuals $\hat{\varepsilon}_i = y_i - \mu(\mathbf{x}_i, \hat{\boldsymbol{\theta}})$ compared to the original y_i 's. If $y_i \doteq \mu(\mathbf{x}_i, \hat{\boldsymbol{\theta}})$, then $V_p(\hat{Y}_{GD}) \doteq 0$, where V_p denotes design-variance. On the other hand, if the relationship between y and \mathbf{x} is not strong enough, we might have $V_p(\hat{Y}_{GD}) > V_p(\hat{Y}_{HT})$, no gain by using \hat{Y}_{GD} . This may be true even when the working-model is correct. The model-calibration estimator \hat{Y}_{MC} , on the other hand, uses $\hat{\mu}(\mathbf{x}_i, \hat{\boldsymbol{\theta}})$'s as a tool of calibration while keeping as close to \hat{Y}_{HT} as possible. It is arguable that \hat{Y}_{MC} will perform much better. The reduction of the design variance in using \hat{Y}_{MC} , as compared to using \hat{Y}_{HT} , is associated with a positive correlation between y_i and $\mu(\mathbf{x}_i, \hat{\boldsymbol{\theta}})$. \hat{Y}_{MC} can be viewed algebraically as a regression estimator using $\mu(\mathbf{x}_i, \hat{\boldsymbol{\theta}})$ as the auxiliary variable. Even in the case of model misspecification, the gain from using a regression estimator on these variables is still available.

3.5 A Comparison Under Simple Random Sampling

To illustrate the points discussed in the previous section, let us make a heuristic comparison under simple random sampling. Let μ_i , $i = 1, \dots, N$, be the values of $\mu(\mathbf{x}_i, \boldsymbol{\theta})$ treated as a single auxiliary variable and ignoring the fact that the parameters $\boldsymbol{\theta}$ have been estimated (in fact, as we see in the proof of Theorem 1, Part 2, the estimated model parameters do not change the asymptotic design variance) and let $v_i = q_i = 1$. Let \bar{U} , σ_U^2 , σ_Y^2 and ρ be the finite population mean and variance of the μ_i 's, the finite population variance of the y_i 's and the finite population correlation between the y_i 's and μ_i 's, respectively. For example, $\sigma_Y^2 = N^{-1} \sum_{i=1}^N (y_i - \bar{Y})^2$. We

can then rewrite \hat{Y}_{GD} and \hat{Y}_{MC} as

$$\hat{Y}_{GD} = \bar{y} - \{\bar{\mu} - \bar{U}\}$$

and

$$\hat{Y}_{MC} = \bar{y} - \{\bar{\mu} - \bar{U}\} \hat{B},$$

where \bar{y} and $\bar{\mu}$ are sample means, $\hat{B} = \hat{\rho} \hat{\sigma}_Y / \hat{\sigma}_U$ and $\hat{\rho}$, $\hat{\sigma}_U^2$ and $\hat{\sigma}_Y^2$ are the sample versions of ρ , σ_U^2 and σ_Y^2 . Letting $B = \rho \sigma_Y / \sigma_U$ we get

$$\hat{Y}_{MC} = \bar{y} - \{\bar{\mu} - \bar{U}\} B + o_p(n^{-1/2}).$$

It is straightforward to show that

$$V_p(\hat{Y}_{GD}) - V_p(\hat{Y}_{MC}) \doteq \frac{1-f}{n} \sigma_Y^2 \left\{ \rho - \frac{\sigma_U}{\sigma_Y} \right\}^2 \geq 0,$$

where V_p denotes design-based variance. Also note that

$$V_p(\hat{Y}_{GD}) \doteq \frac{1-f}{n} \{ \sigma_Y^2 + \sigma_U^2 - 2\sigma_{YU}^2 \},$$

in obvious notation, and thus if $\sigma_U^2 > 2\sigma_{YU}^2$, then $V_p(\hat{Y}_{GD}) > V_p(\bar{y}) = V_p(\hat{Y}_{HT})$. That is, \hat{Y}_{GD} will perform worse than \hat{Y}_{HT} if $\rho^2 < 1/(2\sigma_Y^2)$.

While for the model-calibration estimator,

$$V_p(\hat{Y}_{MC}) \doteq \frac{1-f}{n} \sigma_Y^2 (1 - \rho^2) \leq V_p(\hat{Y}_{HT}),$$

the gain from using \hat{Y}_{MC} is almost guaranteed. These relationships are verified in the limited simulation study in Section 3.6.

3.6 A Simulation

We conducted a limited simulation study to investigate the finite sample performance of the estimators of \bar{Y} proposed in Sections 3.2–3.4. A finite population consisting of $N = 2,000$ units was generated as an iid sample from $\log(y) = \theta_0 + \theta_1 x + \varepsilon$, where $x \sim \text{Gamma}(1, 1)$ and $\varepsilon \sim N(0, \sigma^2)$. We chose $\theta_0 = \theta_1 = 1$. Four different finite populations were used by choosing different values of σ^2 such that the correlation coefficient between $\log(y)$ and x are .9, .8, .7, and .6, respectively.

For each fixed finite population, a simple random sample of size $n = 100$ was taken and a log-linear model

$$\log(\mu_i) = \alpha + \beta x_i, \quad v(\mu) = \mu^2$$

was fit using pseudomaximum quasilielihood estimation. Estimators \hat{Y}_{MC} , \hat{Y}_{MC}^* , \hat{Y}_{EL} and \hat{Y}_{GD} were computed using the sample data and all the fitted values. We also included the GREG based on a linear model in the simulation to compare to a routine application without modeling. All estimators were compared to the baseline estimator, \hat{Y}_{HT} . The process was repeated $B = 50,000$ times.

Table 1. Relative Bias (RB%) and Relative Efficiency (RE) of Estimators for Estimating the Finite Population Mean

Population	ρ	\hat{Y}_{HT}	\hat{Y}_{MC}	\hat{Y}_{MC}^*	\hat{Y}_{EL}	\hat{Y}_{GD}	\hat{Y}_{GR}
Percentage Relative Bias (RB)							
1	.9	.15	-1.23	-1.21	-1.12	-.22	-4.97
2	.8	.14	-2.10	-2.14	-2.09	.18	-5.10
3	.7	.07	-2.78	-2.95	-3.05	1.29	-5.27
4	.6	-.33	-2.87	-3.32	-3.60	5.13	-5.71
Relative Efficiency to \hat{Y}_{HT} (RE)							
1	.9	1.00	3.33	3.33	3.13	2.38	1.19
2	.8	1.00	2.50	2.44	2.27	1.72	1.15
3	.7	1.00	2.04	2.04	1.85	1.41	1.12
4	.6	1.00	1.64	1.64	1.12	.89	1.11

The performance of the various estimators was measured by the simulated relative bias (RB, in percentage) and relative efficiency (RE), defined by

$$RB = 100 \times B^{-1} \sum_{i=1}^B (\hat{Y} - \bar{Y}) / \bar{Y},$$

$$RE = \sqrt{MSE_{HT}} / \sqrt{MSE}, \tag{18}$$

where $MSE = B^{-1} \sum_{i=1}^B (\hat{Y} - \bar{Y})^2$ and MSE_{HT} is the MSE of \hat{Y}_{HT} .

Table 1 reports RB and RE for the estimators included in the simulation. Several interesting points are highlighted here:

- (1) the RB are all within a reasonable range, with the GREG having the largest at 5%;
- (2) \hat{Y}_{MC} , \hat{Y}_{MC}^* and \hat{Y}_{EL} perform similarly and better in all cases;
- (3) \hat{Y}_{MC}^* never outperforms \hat{Y}_{MC} , and \hat{Y}_{EL} never outperforms \hat{Y}_{MC} or \hat{Y}_{MC}^* . The reason for the latter may represent the price to be paid to achieve the positive weights;
- (4) \hat{Y}_{GD} performs well when the relationship between y and x is strong (populations 1 and 2), but can be worse than \hat{Y}_{HT} , which does not even use the auxiliary information, when the relationship is weak (population 4);
- (5) the gain from using the GREG, which ignores the curved relationship between y and x , is always marginal.

4. ESTIMATING THE FINITE POPULATION DISTRIBUTION FUNCTION

The finite population distribution function evaluated at t is defined as the proportion of units with y values less than or equal to t ,

$$F_N(t) = N^{-1} \sum_{i=1}^N I_{[y_i \leq t]},$$

where $I_{[\cdot]}$ is the indicator function. By replacing y_i by $I_{[y_i \leq t]}$, many of the estimators that were constructed for estimating the population mean can be used for estimating $F_N(t)$. For instance, the Horvitz–Thompson estimator for $F_N(t)$ is $\hat{F}_{HT}(t) = N^{-1} \sum_{i \in s} d_i I_{[y_i \leq t]}$, which makes no use of auxiliary information at the estimation stage. However, such a transplantation is not always workable when auxiliary information

is to be used. Part of the reason is that, for example, a simple linear regression model assumed for y and x cannot be transmitted to $I_{[y \leq t]}$ and x or $I_{[y \leq t]}$ and $I_{[x \leq t]}$. The model must be used in its original form as we deal with the dichotomous variable $I_{[y \leq t]}$.

Our approaches proposed in Section 3 for the mean case also provide a unified framework for the estimation of distribution functions. For the distribution function, it is the fitted values of $I_{[y_i \leq t]}$, not of y_i , that should be used.

4.1 Estimation of $F_N(t)$ Under a Regression Model

Suppose the superpopulation follows the regression model given in (5): $y_i = \mu(\mathbf{x}_i, \boldsymbol{\theta}) + v(\mathbf{x}_i)\varepsilon_i$, $i = 1, 2, \dots, N$. Letting $G(\cdot)$ be the cumulative distribution function of the error term ε_i implies

$$E_{\varepsilon}(I_{[y_i \leq t]} | \mathbf{x}_i) = G\{(t - \mu(\mathbf{x}_i, \boldsymbol{\theta})) / v(\mathbf{x}_i)\}.$$

We can estimate $G(u)$ by $\hat{G}_n(u) = n^{-1} \sum_{i \in s} I_{[\hat{\varepsilon}_i \leq u]}$, where $\hat{\varepsilon}_i = \{y_i - \mu(\mathbf{x}_i, \hat{\boldsymbol{\theta}})\} / v(\mathbf{x}_i)$ and then let $\hat{G}_i = \hat{G}_n\{(t - \mu(\mathbf{x}_i, \hat{\boldsymbol{\theta}})) / v(\mathbf{x}_i)\}$. We suggest using a uniform q_i in the model-calibration estimator of $F_N(t)$. Replacing y_i by $I_{[y_i \leq t]}$ and $\mu(\mathbf{x}_i, \hat{\boldsymbol{\theta}})$ by \hat{G}_i in (9), (15), and (17) yield

- 1. The model-calibration estimator of $F_N(t)$,

$$\hat{F}_{MC}(t) = N^{-1} \sum_{i \in s} d_i I_{[y_i \leq t]} + \left\{ N^{-1} \sum_{i=1}^N \hat{G}_i - N^{-1} \sum_{i \in s} d_i \hat{G}_i \right\} \hat{B}, \tag{19}$$

where $\hat{B} = \sum_{i \in s} d_i (\hat{G}_i - \bar{G})(I_{[y_i \leq t]} - \bar{I}) / \sum_{i \in s} d_i (\hat{G}_i - \bar{G})^2$, $\bar{G} = \sum_{i \in s} d_i \hat{G}_i / \sum_{i \in s} d_i$, $\bar{I} = \sum_{i \in s} d_i I_{[y_i \leq t]} / \sum_{i \in s} d_i$;

- 2. The pseudoempirical maximum likelihood estimator of $F_N(t)$,

$$\hat{F}_{EL}(t) = \sum_{i \in s} \hat{p}_i I_{[y_i \leq t]}, \tag{20}$$

where \hat{p}_i 's maximize the pseudoempirical likelihood $\hat{l}(\mathbf{p})$, subject to

$$\sum_{i \in s} p_i = 1, \quad \sum_{i \in s} p_i \hat{G}_i = N^{-1} \sum_{i=1}^N \hat{G}_i \quad (0 \leq p_i \leq 1); \tag{21}$$

- 3. The model-assisted difference estimator of $F_N(t)$,

$$\hat{F}_{GD}(t) = N^{-1} \left\{ \sum_{i \in s} d_i I_{[y_i \leq t]} - \sum_{i \in s} d_i \hat{G}_i + \sum_{i=1}^N \hat{G}_i \right\}. \tag{22}$$

At this point, we will make a few comments on these three estimators. First, $\hat{F}_{GD}(t)$ is essentially the estimator proposed by Rao, Kovar, and Mantel (1990) under model (5). This same estimator was also derived by Godambe (1989) using the model- and design-based optimum estimating function theory. Next, note that \hat{p}_i 's used here for $\hat{F}_{EL}(t)$ are not the same as those used in \hat{Y}_{EL} , where the constraint (15) is imposed on $\hat{\mu}_i$ directly. This is because of the second constraint in (21)

which restricts the predicted G_i 's and not the predicted y_i 's. Also, it is important to note that none of the three estimators will necessarily be monotonic. However, we can say that $0 \leq \hat{F}_{EL}(t) \leq 1$, which may not always be the case for $\hat{F}_{MC}(t)$ or $\hat{F}_{GD}(t)$. It may not be immediately obvious that $\hat{F}_{EL}(t)$ can be nonmonotonic unless one notes the implicit dependence of the \hat{p}_i 's on t .

Asymptotic design-variance and variance estimation for the distribution function can be developed in a similar fashion as for the mean case. However, the crucial step used to establish the asymptotic variance formula for the mean case, Equation (28) in the Appendix, which implies that the estimated model parameters will not change the asymptotic design-variance, is not readily available here because a Taylor series expansion can not be applied to the distribution function. Indeed, such a statement is not available for a general sampling design. Wu and Sitter (2000) established this for some commonly used designs including stratified multistage sampling,

$$\frac{1}{N} \sum_{i=1}^N \hat{G}_i - \frac{1}{N} \sum_{i \in s} d_i \hat{G}_i = \frac{1}{N} \sum_{i=1}^N G_i - \frac{1}{N} \sum_{i \in s} d_i G_i + o_p\left(\frac{1}{\sqrt{n}}\right),$$

(see also, Rao, Kovar, and Mantel 1990). Hence the asymptotic design-variance, $V[\hat{F}_{MC}(t)]$, is given by (11) with U_i replaced by $I_{[y_i \leq t]} - G_i B$ and the variance estimator is given by (12) with u_i replaced by $I_{[y_i \leq t]} - \hat{G}_i \hat{B}$. Here, B is the finite population quantity corresponding to \hat{B} . We should also note that, the same variance estimator of \hat{F}_{MC} can be used for \hat{F}_{EL} and a variance estimator for \hat{F}_{GD} is obtained by letting $\hat{B} = 1$ in $v(\hat{F}_{MC})$.

The question of how to do quantile estimation is a natural one at this point. This is discussed for $\hat{F}_{GD}(t)$ by Rao, Kovar, and Mantel (1990). They suggest first making $\hat{F}_{GD}(t)$ monotone nondecreasing via the method of Francisco and Fuller (1991), and then inverting it. A similar approach may be used for all three of the estimators described. Confidence intervals and variances estimates for quantiles can then be obtained using Woodruff's method (see Francisco and Fuller 1991; Kovar, Rao, and Wu 1988; Sitter and Wu in press, and Woodruff 1952). We do not pursue this further in this article, intending to consider possibly more elegant approaches in future investigations.

4.2 Estimation of $F_N(t)$ Under a General Model

Note that none of the estimators in Section 4.1 satisfy the desirable property that

$$\hat{F}_N(t) = F_N(t) \quad \text{if } y_i = \mu(\mathbf{x}_i, \boldsymbol{\theta}), \quad i = 1, 2, \dots, N, \quad (23)$$

and the construction of these estimators relies on the error distribution function, $G(\cdot)$, which is associated with the regression model. To achieve (23) under a general modeling process, we similarly define $\hat{F}_{MC}^*(t)$, $\hat{F}_{EL}^*(t)$ and $\hat{F}_{GD}^*(t)$ as in Section 4.1, but replace \hat{G}_i by $I_{[\mu(\mathbf{x}_i, \hat{\boldsymbol{\theta}}) \leq t]}$ in (19), (20), and (22). The so-defined $\hat{F}_{GD}^*(t)$ was also introduced by Rao, Kovar, and Mantel (1990) under the context of a linear regression working model. All these modified estimators satisfy $\hat{F}_N(t) \doteq F_N(t)$ if $y_i = \mu(\mathbf{x}_i, \boldsymbol{\theta})$, $i = 1, 2, \dots, N$. They are also

Table 2. Simulated Relative Efficiency to $\hat{F}_{HT}(t)$ for Estimating the Distribution Function $F_N(t)$ at $t = t_p$ Such That $F_N(t) = p$

Population	p	p	$\hat{F}_{HT}(t_p)$	$\hat{F}_{MC}^*(t_p)$	$\hat{F}_{EL}^*(t_p)$	$\hat{F}_{GD}^*(t_p)$
1	.9	.10	1.00	1.00	1.00	1.00
		.25	1.00	1.43	1.05	.97
		.50	1.00	1.30	1.30	1.19
		.75	1.00	1.45	1.41	1.37
		.90	1.00	1.67	1.37	1.61
2	.8	.10	1.00	1.00	1.00	1.00
		.25	1.00	1.00	1.01	.99
		.50	1.00	1.14	1.12	.98
		.75	1.00	1.25	1.24	1.14
		.90	1.00	1.35	1.24	1.27
3	.7	.10	1.00	1.00	1.00	1.00
		.25	1.00	1.00	1.00	1.00
		.50	1.00	1.05	1.04	.90
		.75	1.00	1.15	1.12	.99
		.90	1.00	1.22	1.16	1.11
4	.6	.10	1.00	1.00	1.00	1.00
		.25	1.00	1.00	1.00	1.00
		.50	1.00	1.01	1.01	.93
		.75	1.00	1.09	1.08	.89
		.90	1.00	1.12	1.10	.96

design-consistent. If the model is correctly specified, $\hat{F}_{MC}^*(t)$, $\hat{F}_{EL}^*(t)$ and $\hat{F}_{GD}^*(t)$ should perform well.

We performed a small simulation to compare these estimators. We use the same simulation setup as in Section 3.6 with a log-linear model relating y to x . We again form a finite population of size $N = 2,000$ and run $B = 50,000$ simulations. The resulting percentage relative biases are all less than 1%, therefore we do not present them so as to save space (see Wu 1999). The relative efficiency (RE) values for $\hat{F}_{MC}^*(t)$, $\hat{F}_{EL}^*(t)$ and $\hat{F}_{GD}^*(t)$, defined as in (18) but replacing \hat{Y} and \bar{Y} by $\hat{F}_N(t)$ and $F_N(t)$ throughout, are presented in Table 2 for $t = t_p$ such that $F_N(t_p) = p$, $p = .1, .25, .5, .75$ and $.9$. We include $\hat{F}_{HT}(t)$ for baseline comparison.

Viewing Table 2, we see that (1) $\hat{F}_{MC}^*(t)$ and $\hat{F}_{EL}^*(t)$ perform similarly and much better than $\hat{F}_{GD}^*(t)$ and $\hat{F}_{HT}(t)$ in almost all cases; (2) $\hat{F}_{GD}^*(t)$ performs slightly worse than $\hat{F}_{HT}(t)$ in some cases; and (3) all three of $\hat{F}_{MC}^*(t)$, $\hat{F}_{EL}^*(t)$ and $\hat{F}_{GD}^*(t)$ show highest efficiency gains at higher percentiles. This last point is really just a property of the particular populations (models) being considered here. Considering Population 2 as plotted in Figure 1, we see that in the low percentile (small y) region, x is a poor predictor of y .

5. CONCLUDING REMARKS

We have proposed a model-calibration approach to the use of complete auxiliary information in complex surveys to estimate totals, means, and distribution functions. The idea involves fitting a general working model and then calibrating on the resulting fitted values as opposed to on the auxiliary variables themselves.

We can summarize the innovation in this work as follows: The relationship between an assumed model and the use of complete auxiliary information is highlighted by noting that, in the case of a linear working-model, it is only necessary to know the mean of the auxiliary variables for the entire finite population to construct efficient estimators of \bar{Y} .

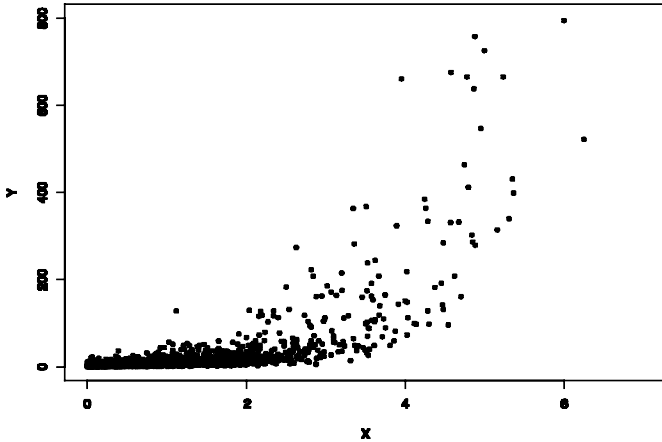


Figure 1. Scatterplot of Population 2.

Therefore, making more effective use of complete auxiliary information requires more complex modeling. Next, the most obvious direction for extending to nonlinear models is through the generalized difference estimator. However, as we demonstrate, both theoretically and through simulation, unless the relationship between y and \mathbf{x} is very strong, this approach can do quite poorly and in fact can perform worse than ignoring the auxiliary information altogether. It is not obvious that one can/should avoid calibrating on the vector of auxiliary variables directly, where a constrained high dimensional maximization or minimization problem needs to be solved. We argue and demonstrate that a simple and powerful way to do this is to calibrate on the fitted values either directly or using a pseudoempirical likelihood approach. Finally, the estimation of the finite population distribution function using complete auxiliary information is shown to be part of this unified approach through the fitted values of the indicator variable $I_{[y \leq t]}$.

APPENDIX: PROOF OF THEOREM 1

For vectors θ , $\theta^{(1)}$ and $\theta^{(2)}$, $\theta = \theta^{(1)} + O_p(1/\sqrt{n})$ or $\theta \in (\theta^{(1)}, \theta^{(2)})$ is to mean $\theta_k = \theta_k^{(1)} + O_p(1/\sqrt{n})$ or $\theta_k \in (\theta_k^{(1)}, \theta_k^{(2)})$ for each component of the vectors; $|\theta|$ represents the usual L_1 norm.

(1) We need to establish the relationships among some population totals defined through $\mu(\mathbf{x}_i, \theta_N)$ and their estimates through $\mu(\mathbf{x}_i, \hat{\theta})$. From (i) and (ii), by applying a Taylor series approximation to $\mu(\mathbf{x}_i, \hat{\theta})$ at $\theta = \theta_N$, we get

$$\mu(\mathbf{x}_i, \hat{\theta}) = \mu(\mathbf{x}_i, \theta_N) + \left\{ \frac{\partial \mu(\mathbf{x}_i, \mathbf{t})}{\partial \mathbf{t}} \Big|_{\mathbf{t}=\theta^*} \right\}' (\hat{\theta} - \theta_N), \quad (24)$$

where $\theta^* \in (\hat{\theta}, \theta_N)$ or $(\theta_N, \hat{\theta})$. Note that, by conditions (i), (ii) and expansion (24),

$$N^{-1} \sum_{i=1}^N \mu(\mathbf{x}_i, \hat{\theta}) = N^{-1} \sum_{i=1}^N \mu(\mathbf{x}_i, \theta_N) + O_p(n^{-1/2}), \quad (25)$$

$$N^{-1} \sum_{i \in s} d_i \mu(\mathbf{x}_i, \hat{\theta}) = N^{-1} \sum_{i \in s} d_i \mu(\mathbf{x}_i, \theta_N) + O_p(n^{-1/2}). \quad (26)$$

Also note that, by (iii), $N^{-1} \sum_{i=1}^N \mu(\mathbf{x}_i, \theta_N) - N^{-1} \sum_{i \in s} d_i \mu(\mathbf{x}_i, \theta_N) = O_p(n^{-1/2})$. This, together with (25) and (26), implies that $N^{-1} \sum_{i=1}^N \mu(\mathbf{x}_i, \hat{\theta}) - N^{-1} \sum_{i \in s} d_i \mu(\mathbf{x}_i, \hat{\theta}) = O_p(n^{-1/2})$. Because $\hat{B}_N = O_p(1)$, asymptotic design unbiasedness of \hat{Y}_{MC} follows from $\hat{Y}_{MC} =$

$\hat{Y}_{HT} + O_p(n^{-1/2})$. The same is true for \hat{Y}_{MC}^* . If $\hat{\theta} \rightarrow \theta$, where θ is the true superpopulation parameter, it then follows that $E_{\xi}(\hat{B}_N) \doteq 1$, $E_{\xi}(\hat{B}_N^*) \doteq 1$, $E_{\xi}(\hat{Y}_{MC} - \bar{Y}) \doteq 0$ and $E_{\xi}(\hat{Y}_{MC}^* - \bar{Y}) \doteq 0$, where expectation is taken under the assumed model.

(2) With the additional condition (iv), a second-order Taylor series approximation to $\mu(\mathbf{x}_i, \theta)$ at $\theta = \theta_N$ is given by

$$\begin{aligned} \mu(\mathbf{x}_i, \hat{\theta}) &= \mu(\mathbf{x}_i, \theta_N) + \left\{ \frac{\partial \mu(\mathbf{x}_i, \mathbf{t})}{\partial \mathbf{t}} \Big|_{\mathbf{t}=\theta_N} \right\}' (\hat{\theta} - \theta_N) \\ &\quad + (\hat{\theta} - \theta_N)' \left\{ \frac{\partial^2 \mu(\mathbf{x}_i, \mathbf{t})}{\partial \mathbf{t} \partial \mathbf{t}'} \Big|_{\mathbf{t}=\theta^*} \right\} (\hat{\theta} - \theta_N), \end{aligned} \quad (27)$$

where $\theta^* \in (\hat{\theta}, \theta_N)$ or $(\theta_N, \hat{\theta})$. It follows from (iv) and (27) that

$$\begin{aligned} N^{-1} \sum_{i=1}^N \mu(\mathbf{x}_i, \hat{\theta}) &= N^{-1} \sum_{i=1}^N \mu(\mathbf{x}_i, \theta_N) + \left\{ N^{-1} \sum_{i=1}^N \mathbf{K}(\mathbf{x}_i, \theta_N) \right\}' \\ &\quad \times (\hat{\theta} - \theta_N) + O_p\left(\frac{1}{n}\right), \\ N^{-1} \sum_{i \in s} d_i \mu(\mathbf{x}_i, \hat{\theta}) &= N^{-1} \sum_{i \in s} d_i \mu(\mathbf{x}_i, \theta_N) + \left\{ N^{-1} \sum_{i \in s} d_i \mathbf{K}(\mathbf{x}_i, \theta_N) \right\}' \\ &\quad \times (\hat{\theta} - \theta_N) + O_p\left(\frac{1}{n}\right), \end{aligned}$$

where $\mathbf{K}(\mathbf{x}_i, \theta_N) = (\partial \mu(\mathbf{x}_i, \mathbf{t}) / \partial \mathbf{t})|_{\mathbf{t}=\theta_N}$. By conditions (i) and (iii), $\hat{\theta} - \theta_N = O_p(1/\sqrt{n})$, $N^{-1} \sum_{i \in s} d_i \mathbf{K}(\mathbf{x}_i, \theta_N) - N^{-1} \sum_{i=1}^N \mathbf{K}(\mathbf{x}_i, \theta_N) = O_p(1/\sqrt{n})$. Therefore,

$$\begin{aligned} N^{-1} \sum_{i=1}^N \hat{\mu}_i - N^{-1} \sum_{i \in s} d_i \hat{\mu}_i &= N^{-1} \sum_{i=1}^N \mu_i - N^{-1} \sum_{i \in s} d_i \mu_i + O_p\left(\frac{1}{n}\right), \end{aligned} \quad (28)$$

where $\hat{\mu}_i = \mu(\mathbf{x}_i, \hat{\theta})$ and $\mu_i = \mu(\mathbf{x}_i, \theta_N)$. Because $\hat{B}_N = B_N + o_p(1)$, we obtain

$$\begin{aligned} \hat{Y}_{MC} &= \hat{Y}_{HT} + \left\{ N^{-1} \sum_{i=1}^N \mu_i - N^{-1} \sum_{i \in s} d_i \mu_i \right\} B_N + o_p\left(\frac{1}{\sqrt{n}}\right) \\ &= N^{-1} \sum_{i \in s} d_i U_i + N^{-1} \sum_{i=1}^N \mu_i B_N + o_p\left(\frac{1}{\sqrt{n}}\right), \end{aligned}$$

where $U_i = y_i - \mu_i B_N$. The asymptotic design variance $V(\hat{Y}_{MC})$ is the same as the design variance of a Horvitz-Thompson estimator $N^{-1} \sum_{i \in s} d_i U_i$, which is given by (11). This variance can be estimated by (12). Similar arguments hold for $V(\hat{Y}_{MC}^*)$.

(3) Consider a more general linear model,

$$\mu(\mathbf{x}_i, \theta) = \theta_0 + \theta_1 u_1(\mathbf{x}_i) + \dots + \theta_p u_p(\mathbf{x}_i),$$

which includes (13) as special case. In this case, $\hat{\mu}_i = \mu(\mathbf{x}_i, \hat{\theta}) = \mathbf{Z}_i' \hat{\theta}$, where $\mathbf{Z}_i = (1, u_1(\mathbf{x}_i), \dots, u_p(\mathbf{x}_i))'$. The pseudomaximum quasi-likelihood estimator and the weighted least squares estimator of θ_N have the same form: $\hat{\theta} = \{\mathbf{Z}' \mathbf{W}^{-1} \mathbf{Z}\}^{-1} \mathbf{Z}' \mathbf{W}^{-1} \mathbf{Y}_n$, where $\mathbf{Z} = (\mathbf{Z}_1, \dots, \mathbf{Z}_n)'$, $\mathbf{W}^{-1} = \text{diag}(d_1 q_1, \dots, d_n q_n)$ and $\mathbf{Y}_n = (y_1, \dots, y_n)'$. It follows that

$$\sum_{i \in s} d_i q_i \hat{\mu}_i^2 = \mathbf{Y}_n' \mathbf{W}^{-1} \mathbf{Z} \{\mathbf{Z}' \mathbf{W}^{-1} \mathbf{Z}\}^{-1} \mathbf{Z}' \mathbf{W}^{-1} \mathbf{Y}_n = \sum_{i \in s} d_i q_i \hat{\mu}_i y_i,$$

and $\hat{B}_N^* = 1$.

Letting $\hat{\boldsymbol{\mu}}_n = (\hat{\boldsymbol{\mu}}_1, \dots, \hat{\boldsymbol{\mu}}_n)' = \mathbf{Z}\hat{\boldsymbol{\theta}}$, it can be shown that $\mathbf{Z}'\{\mathbf{W}^{-1}\mathbf{Y}_n - \mathbf{W}^{-1}\hat{\boldsymbol{\mu}}_n\} = 0$. Since the first column of \mathbf{Z} is a vector of 1's, we get $\sum_{i \in s} d_i q_i y_i = \sum_{i \in s} d_i q_i \hat{\boldsymbol{\mu}}_i$, and $\hat{B}_N = 1$. Thus $\hat{Y}_{MC} = \hat{Y}_{MC}^* = \hat{Y}_{GD} = \hat{Y}_{GR}$.

[Received February 1999. Revised December 1999.]

REFERENCES

- Basu, D. (1971), *Foundations of Statistical Inference, A Symposium*, eds. V. P. Godambe and D. A. Sprott, Toronto: Holt Rinehart and Winston.
- Binder, D. A. (1983), "On the Variances of Asymptotically Normal Estimators from Complex Surveys," *International Statistical Review*, 51, 279–292.
- Cassel, C. M., Särndal, C. E., and Wretman, J. H. (1976), "Some Results on Generalized Difference Estimation and Generalized Regression Estimation for Finite Populations," *Biometrika*, 63, 615–620.
- Chen, J., and Qin, J. (1993), "Empirical Likelihood Estimation for Finite Populations and the Effective Usage of Auxiliary Information," *Biometrika*, 80, 107–116.
- Chen, J., and Sitter, R. R. (1999), "A Pseudo Empirical Likelihood Approach to the Effective Use of Auxiliary Information in Complex Surveys," *Statistica Sinica*, 9, 385–406.
- Deville, J. C., and Särndal, C. E. (1992), "Calibration Estimators in Survey Sampling," *Journal of the American Statistical Association*, 87, 376–382.
- Francisco, C. A., and Fuller, W. A. (1991), "Quantiles Estimation With a Complex Survey Design," *The Annals of Statistics*, 19, 454–469.
- Godambe, V. P. (1989), "Estimation of Cumulative Distribution of a Survey Population," Tech. Report No. STAT-89-17. Department of Statistics and Actuarial Science, University of Waterloo.
- Godambe, V. P., and Thompson, M. E. (1986), "Parameters of Superpopulation and Survey Population: Their Relationships and Estimation," *International Statistical Review*, 54, 127–138.
- Kovar, J. G., Rao, J. N. K., and Wu, C. F. J. (1988), "Bootstrap and Other Methods to Measure Errors in Survey Estimates," *The Canadian Journal of Statistics*, 16, Supplement, 25–45.
- Molina, C. E. A., and Skinner, C. J. (1992), "Pseudo-likelihood and Quasi-likelihood Estimation for Complex Sampling Schemes," *Computational Statistics and Data Analysis*, 13, 395–405.
- Rao, J. N. K. (1966), "Alternative Estimators in PPS Sampling for Multiple Characteristics," *Sankhya*, Ser. A, 28, 47–60.
- 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.
- Särndal, C. E., (1980), "On π -inverse Weighting Versus Best Linear Unbiased Weighting in Probability Sampling," *Biometrika*, 67, 639–650.
- Sitter, R. R. and Wu, C. (in press), "A Note on Woodruff Confidence Intervals for Quantiles," *Statist. Prob. Letters*.
- Woodruff, R. S. (1952), "Confidence Intervals for Medians and Other Position Measures," *Journal of the American Statistical Association*, 47, 635–646.
- Wu, C. (1999), "The Effective Use of Complete Auxiliary Information From Survey Data," Unpublished doctoral dissertation, Simon Fraser University, Canada.
- Wu, C., and Sitter, R. R. (in press), "Variance Estimation for the Finite Population Distribution Function with Complete Auxiliary Information," *The Canadian Journal of Statistics*, 29.
- Zhong, C. X. B., and Rao, J. N. K. (1996), "Empirical Likelihood Inference Under Stratified Random Sampling Using Auxiliary Information," in *Proceedings of the Section on Survey Research Methods, American Statistical Association*, pp. 798–803.